

Summary of Measurement Intercomparisons During TRACE-P

F.L. Eisele,^{1,2} L. Mauldin,¹ C. Cantrell,¹ M. Zondlo,¹ E. Apel,¹ A. Fried,¹ J. Walega,¹ R. Shetter,¹ B. Lefer,¹ F. Flocke,¹ A. Weinheimer,¹ M. Avery,³ S. Vay,³ G. Sachse,³ H. B. Singh,⁹ W. Brune,⁴ H. Harder,⁴ M. Martinez,⁴ A. Bandy,⁵ D. Thornton,⁵ B. Heikes,⁶ Y. Kondo,⁷ D. Riemer,⁸ S. Sandholm,² D. Tan,² R. Talbot,¹⁰ Jack Dibb¹⁰

1 National Center for Atmospheric Research, Boulder, CO

2 Georgia Institute of Technology, Atlanta, GA

3 NASA Langley Research Center

4 Penn State University, University Park, PA

5 Drexel University, Philadelphia, PA

6 University of Rhode Island

7 University of Tokyo, Tokyo, Japan

8 University of Miami, Miami, Florida

9 NASA Ames Research Center

10 University of New Hampshire, Durham, NH

Abstract

The NASA DC-8 and P-3B aircraft flew within about a km or less of each other on three occasions during the TRACE-P campaign in order to intercompare similar measurements on the two aircraft. The first and last intercomparisons were in relatively remote marine environments with a boundary layer measurement followed by an ascent to 3 km for the first and a measurement at 5.3 km followed by a decent and boundary layer leg for the third. The second set of intercomparisons was at a fixed altitude of about 5.2 km off the coast of Japan, also in relatively clean air. A number of measurements were compared with the best agreement observed for the most abundant compounds such as CO₂ and CH₄ and with very good agreement for CO, O₃ and j values. Other comparisons including measurements of the same compounds on both aircraft and measurements of the same compound by two different instruments on the DC-8 varied over a wide range from quite good for PAN, NO, HNO₃ to reasonable agreement for OH, HO₂, CH₂O, acetone, and methylethylketone and to generally poor for NO₂, SO₂, PPN, acetaldehyde, and methanol. The comparison results particularly those for the first one second CO and O₃ measurements suggest that credible intercomparisons can be made using two aircraft in close proximity. Comparisons such as these, made as a component of a larger field campaign have the advantage that they test the actual instrument configuration used during the field study and they require no additional instrument installation and testing.

Introduction

The intercomparison of many different measurement techniques for a variety of important atmospheric compounds during a field campaign can provide much needed insight into how well various measurements are being made under real world conditions. The importance of comparisons has long been recognized by the NASA/GTE program, which has played a leadership role in designing and sponsoring several such studies [Hoell *et al.*, 1993, Hoell *et al.*, 1990, Beck *et al.*, 1987]. These have, in general, been formal blind or double blind intercomparisons which have generally been separate from a research campaign. The present intercomparison differs in several ways from previous comparisons, but also tries to address new and ever more complex measurement and validation concerns. Former NASA intercomparisons were designed to evaluate which instrument or measurement technique provided the most accurate, precise, and sensitive measurement. An isolated comparison of instruments, of course, does not by itself provide this information, but in conjunction with the use of common calibration standards, ancillary measurements, modeling, and a wide range of natural variability of the compound of interest, much can be learned about relative sensitivity, precision, and to a lesser extent, absolute accuracy. Such intercomparisons were fairly competitive, and at least to some extent, aimed at identifying the best instrument to be used in future NASA missions. While this need still exists, the campaign scene has become more complex, requiring additional comparison opportunities. The NASA GTE program now often flies two aircraft simultaneously during a field campaign, each with its own specialty, that complements the mission goals, but still with many of the same measurement capabilities on each aircraft. In addition, TRACE-P and future missions are planned to be joint with

aircraft from other agencies and with satellites which also possess overlapping measurement capabilities. If models are to combine data from multiple aircraft and even multiple agencies' aircraft and satellites in a meaningful way, biases in measurements between platforms must first be identified and, in the future, minimized or removed. As the number of compounds being measured expands and in many cases the techniques being used to study each of them also expands, with each technique often having its own inherent advantages and disadvantages for a given mission or platform, the problem of intercomparison becomes more difficult. In many cases, there may be no best measurement technique for a given compound or set of compounds. If several different measurements of the same compound can be compared during a field campaign, a diverse set of techniques is probably even preferable, because if they all agree (using very different measurement schemes) the combined set of measurements as a unit becomes far less susceptible to interferences and, in some cases, calibration errors.

There are a growing number of measurements that have very special inlet needs because of surface interactions and/or air speed and altitude dependent sampling. Intercomparison of measurements of such compounds and particles requires that the exact inlet, sample line, and instrument configuration to be used in a field campaign be used for an intercomparison study. It also follows that changes and improvements of sampling technique may invalidate intercomparison results. There is therefore a need to intercompare every new measurement configuration used in a field campaign. Thus, the goal of present and future intercomparisons may be less to identify the best measurement technique and more to evaluate biases between instruments and platforms on a campaign-by-campaign basis, initially correcting for these biases and in future campaigns

minimizing and/or eliminating them. In some cases, this may require discontinuing the use of certain techniques which are (a) inconsistent with other measurements, model predictions, and which can not be verified by independent means, or (b) those that are too insensitive or slow to answer the questions they are meant to address. In most cases, however, it means intercomparing calibration standards, identifying interferences, finding inlet losses or enhancements, and determining the real world range of altitudes, speeds, temperatures, humidities, etc. over which measurements can consistently be made within some predetermined error limits.

The present intercomparison is the first attempt within a GTE mission to address these broader issues. The intercomparison was completely informal with open data sharing throughout the study period. It involved three .5-1.5 hour comparison periods which were part of a larger field campaign. The advantage of this approach is that it compares instruments in the same configuration in which they are used during the mission and helps to identify instrument malfunctions during the field campaign. Comparisons can be made under a range of conditions typical of those encountered in the measurement campaign. There is no additional setup time or cost for installing instrumentation on the aircraft beyond that already required for the field campaign. The limitations of this approach are that only a short period of time is available to conduct the comparison portion of the study (this is a much larger problem for long integration time measurements). There is also some uncertainty that the two aircraft are sampling the same air mass throughout the intercomparison period, and even when they are, the time that various features are sampled may vary by a few seconds from aircraft to aircraft. The advent of several one second chemical measurements and the large number of

simultaneous measurements on the two aircraft, however, has greatly reduced, but certainly not eliminated the latter concern for close aircraft proximity as will be discussed shortly. The present intercomparison was also not blind, but rather encouraged full sharing of data to address any concerns that arose. Considering that this was the first attempt at this type of intercomparison and very much a learning experience, it was very reasonable that the study was not blind. There is nothing, however, that prevents this same strategy from being applied to a blind study in the future by collecting data submitted in a blind fashion shortly after an intercomparison flight and then disclosing the data after its formal collection to all interested in its use for science planning and model intercomparisons.

If measurements on two aircraft as different as the P-3B and the DC-8 can be successfully intercompared over a wide range of altitudes, then it would appear that intercomparison with most other research aircraft would also be possible in the future. A second successful set of intercomparisons were carried out between the NASA P-3B and the NCAR C-130 near the end of TRACE-P and the beginning of ACE-Asia (Aerosol Characterization Experiment-off the coast of Asia) . These intercomparisons involved largely aerosol instrumentation which are discussed in detail by *K. Moore et al.* (this issue) and *Weber et al.* (this issue). A very successful intercomparison of SO₂ instruments on the same two aircraft is discussed Thornton et al (2002).

The results of the present intercomparisons provide insight into how to combine measurements from the two aircraft involved in TRACE-P into a single merged data set. In some cases, measurements from the two aircraft are essentially indistinguishable, while in others there are distinct differences that need to be acknowledged and possibly

even compensated for in model comparisons. In the non-competitive spirit in which this intercomparison was performed and because there are still concerns about exactly how similar sampled air masses were, persistent differences and trends in data were pointed out, along with detection limit problems and time resolution concerns but individual techniques were not critically reviewed. The details of individual measurements are contained in the many companion papers in this issue.

The purpose of this manuscript is only to provide a summary of possible biases when combining data from various measurement techniques, provide additional information to evaluate the uncertainty or lack thereof that might be encountered in the use of this data, and present a brief analysis of pitfalls and benefits that can be derived from future mission-based intercomparison studies. It is also hoped that the present summary of results will encourage open exchange to foster a better understanding of discrepancies and not focus counter productive emphasis on value judgments particularly for this very informal and somewhat experimental comparison.

Intercomparison Details

The DC-8 and P-3B flew in close proximity to each other on three occasions for the purpose of intercomparing similar measurements on both platforms during the TRACE-P mission. The intercomparisons varied in length from slightly shorter than a half hour to a little longer than an hour and a quarter, and over an altitude range of about 0.16 to 5.3 km. All three of these intercomparisons were conducted in fairly unpolluted air masses. The first was conducted in the boundary layer and during a climb up to 3 km from about 14° latitude and 140-143° longitude on a transit flight from Guam to Hong Kong. The second was conducted at a fixed altitude of about 5.2 km off the coast of

Japan at about 33° latitude and from 137.5 - 141° longitude. The final and by far the longest intercomparison began with a fixed altitude flight at 5.3 km, then gradually descended into the boundary layer for a fixed altitude flight at 0.2 km. These flights covered a latitude range from 22.5° - 25° N and a longitude range from 208° - 214° on the return transit flight out of Hawaii. Figure 1 shows the altitude of both aircraft throughout the three intercomparison periods. It also shows the approximate distance between the two aircraft, which may be uncertain by about 0.1 km. The intercomparison periods are largely defined by aircraft altitude, since the largest chemical differences would be expected to occur in the vertical direction. Also, differences in the horizontal direction along the flight path will show up as changes as a function of time with 0.1 km corresponding to less than a second time shift. Of course, the flight paths of the two aircraft will not necessarily intersect all of the same air masses, and even when they do, they will intersect the transition region between air parcels at a random angle. So, a time period equal to or even several times larger than that required to travel a distance equal to the aircraft spacing might be required to reach the same air mass. This is still, however, only on the order of seconds to a few tens of seconds in the extreme case, and most of the instruments being compared acquire data at a similar or slower rate. Most of the comparison data to be discussed will be from the 1 minute merge files. While uncertainties in the vertical direction should be much smaller than in the horizontal direction, differences of tens of meters are still quite possible. Therefore, no attempt will be made to adjust measurements based on altitude or position. As will be observed in the following section, most of the discrepancies that will be discussed correspond to differences over a significant portion of an intercomparison flight or, in many cases, the

entire flight and do not appear to be related to small differences in the time or altitude that an aircraft intercepted an air mass change. There is, however, one 2 or 3 minute long exception in which both NO and OH varied in a consistent manner on each aircraft, but which differed between aircraft by a factor of 2-3. This was also a period in which observations on both aircraft showed a rapid change in NO (and only small changes in CO and O₃) probably indicating that the aircraft were traversing a relatively recently emitted plume. Since these measurements were in the boundary layer with the aircraft a few km apart, it is quite possible that a ship plume or other local pollution source might have influenced the two aircraft somewhat differently. Since the large NO discrepancies only persisted for about 3 one minute data points, the data from 1:23-1:25 during the first intercomparison flight has been removed. These are the only data removed for the 3 intercomparison flights because they show the only obvious air mass difference. This is not to say that the remainder of the intercomparison was flown in identical and uniform air parcels. There were certainly lesser variations that were presumably encountered throughout the flights; however, as shown in the next section, such brief differences will probably not significantly change the comparison result. Also the agreement between several of the very rapid chemical measurements, particularly O₃ and CO, add greatly to the credibility of this two aircraft intercomparison.

A second type of intercomparison was also possible. A few compounds were measured on the DC-8 by two different instruments, including formaldehyde and several oxygenated hydrocarbons. The measurements of these compounds can thus be compared throughout the entire mission, and since they are all on the same platform, there are no air mass similarity issues. In the one minute data set, there are a few points out of the whole

data set (the points around 4:45 on March 21, 2001) that have been removed because they are several times higher than all other measurements for several oxygenated hydrocarbons. If left in, these points would require rescaling figures, impose an unrealistic bias on fitting routines, and would, for several compounds, represent by far the largest single absolute data discrepancy. Also, the meaning of a single rapid plume crossing event by two instruments with very different cycle times, one as long as 180 seconds, is very limited.

The above timing issue is not isolated to dramatic plume crossings, but is ubiquitous throughout the intercomparisons; it simply gets worse in large abrupt gradients. The major problems are differences in integration times. For example; PAN measurements on the P-3B with an integration time of 1-2 seconds measured once each 150 seconds are compared to 120 second integration time measurements on the DC-8. This could easily result in significant measurement difference purely due to timing (location) differences. In fact, PAN measurements compared quite well, but this is probably because the two aircraft comparisons were generally conducted in remote and relatively uniform air masses. This situation also existed for measurements involved in the DC-8 only comparisons which were conducted throughout TRACE-P and included many rapid air mass changes. An example of such measurements are: Singh's methanol data being reported as a 180 second integration typically every 7 minutes, while Apel reports data integration times from 6 to 100 seconds with times between samples typically 240 to 350 seconds or longer. In the latter case, discrepancies were much larger and it is not clear how much of the observed difference is due to measurement timing (location) and how much is due to instrumental measurement differences and response rates to transients.

This concern will be discussed in more detail by Apel and Riemer (this issue). There is also a concern that some instruments are not run on a consistent time base with other instruments. This added some additional uncertainty when comparing results but is an area where improvements can be made in future campaigns.

Two Aircraft Intercomparison Results

The results of this intercomparison can be divided into 4 categories, beginning with measurements of parameters that typically agreed with each other within a percent or two. These are typically measurements of long lived and relatively abundant compounds or photolysis frequencies. A second group of measurements commonly agreed with each other at about the 10% level, and generally within quoted error limits of the measurement. The third group of measurements also typically agreed with each other within the quoted error limits, but their uncertainties were sufficiently large that model comparisons using these measurements should acknowledge the differences observed on the 2 different aircraft platforms. This group contains only OH and HO₂/RO₂, compounds with very short atmospheric lifetimes and low concentrations. Finally, there were several measurements that did not agree within their quoted error limits and commonly disagreed by a factor of 2 or more.

The measurements that fall into the very good agreement category include those of photolysis frequencies and O₃, CO, CO₂ and CH₄ concentrations. All of these measurements do, however, have several elements in common. In each case, the values measured during the two aircraft intercomparison periods were far above the detection limits of the instruments, the instrument and measurement techniques for a given

parameter were essentially the same on both aircraft, and the same PI was responsible for similar measurements on both aircraft. The latter two circumstances were not the case for measurements in the subsequent categories.

Figures 2 and 3 show comparisons between photolysis frequency measurements determined by actinic flux spectroradiometry [Shetter et al., 2002] on the P-3B vs those on the DC-8 for $j(\text{O}^1\text{D})$ and $j(\text{NO}_2)$ respectively. Agreement is typically well within the stated uncertainty ($\pm 10\%$ and $\pm 8\%$ respectively) of these measurements. Some of the minor scatter that is observed for example in $j\text{O}_3$ around $10\text{--}15 \times 10^{-6}$, $45\text{--}50 \times 10^{-6}$ and one point at about 70×10^{-6} in the O_3 plot appears to be caused by transience in $j\text{O}_3$ probably due to clouds that would not necessarily be expected to be measured in a similar manner on both aircraft. These are particularly prevalent in the first and during the last half of the third intercomparison. Unlike the chemical measurements that will be discussed throughout the remainder of this paper, observed similarities in local chemical fluctuations provide little insight into variations in j values caused by more distant clouds. Fortunately, the observed agreement is quite good despite fluctuations in j , and the primary lesson from these figures is that it is preferable not to intercompare measurements in areas of broken cloudiness above or below the aircraft. Along with the scatter plot of data, two other lines are included. The dashed line is a bivariate fit to the data which is weighted using the relative uncertainties given in the TRACE-P data archive. The solid line uses the same weighting but is forced through the origin. Use of the latter provides important insight because essentially all TRACE-P measurements have some means of obtaining a zero measurement value, and thus the data would be expected to converge to a line through the origin. This line is particularly useful for determining

the slope of data that is taken over a small dynamic range far from the origin. More caution should typically be exercised in using the slope of the dashed line, particularly in figures where the origin is not even shown. When data extends over a relatively wide dynamic range, however, significant deviation in slope between the two lines and a large intercept, may indicate potential measurement nonlinearity, an interference, or an unrealistic background measurement. In the above figures, both lines have a slope so close to 1.0 compared to the stated uncertainty that they provide little additional insight. These two photolysis frequencies are shown as a sample of a much larger number of derived j values also with accuracies in the 8-10% range. Agreement is at a similar level for these other photolysis frequencies and thus an order of magnitude more figures would add little additional information.

Figures 4 and 5 show measurements of CH_4 and CO_2 on the P-3B versus the DC-8 by *Bartlett et al.* (this issue) and *Vay et al.* (this issue) respectively. In both cases the agreement is extremely good with slopes equal to 1 to within better than 0.1% (stated uncertainties are $\pm 1\%$ and $\pm 0.25\text{ppmv}$ for CH_4 and CO_2 respectively). Again the dashed line represents an unconstrained bivariate fit which is included in these figures only for consistency. It has much less meaning for data over such a narrow dynamic range and so far from zero. These 2 figures show the best fit of all the data intercompared and also include the 2 most abundant (ppmv range) and longest-lived species intercompared. Extremely good comparisons were also obtained for O_3 measured by *Avery et al.* (this issue) and CO measured by the Sachse group which are shown in Figures 6 and 7 respectively. Both had slopes within 1% of the 1 to 1 line and essentially all scatter was within the $\pm 5\%$ and $\pm 2\%$ accuracy quoted for O_3 and CO respectively. The bivariate fit

slope of 1.096 may be, in part, explained by the P-3B and DC-8 sampling slightly different air masses towards the end of that intercomparison period. This possibility is discussed below. These compounds have a shorter atmospheric lifetime than CH_4 and CO_2 , on the order of days to months rather than years and variations of a factor of 2 in concentration can commonly be observed in adjacent air masses typically due to nonuniform mixing down wind of enhanced source regions. Both of these measurements also provide data at 1 Hz which because of their high precision and the degree of agreement can be used to better understand the relation between the air masses in which each aircraft was flying. Figure 8 shows a plot of O_3 measured on both aircraft as a function of time for a time starting just before the two aircraft came into close proximity and extending throughout the third intercomparison period. Note that as the two aircraft approach each other the air masses in which they are flying have fairly different O_3 concentrations and that once together (~6:04pm) they both observe very similar O_3 and very similar structure in O_3 concentrations as they descend through several very different layers. An equally impressive demonstration of rapid time response and high precision is observed using the measurement of CO in Fig. 9. These impressive figures provide evidence that both aircraft are sampling from a fairly similar part of the same air mass. However, notice that after ~6:50 pm the O_3 measurements on the two aircraft diverge slightly (~1 to 2 ppb). Interestingly, over this same time period, a similar divergence in the CO values (~2 to 4 ppb) can be seen between the aircraft. These slightly different airmasses (~2-4 ppb CO difference) affect the CO values around 150 ppb at the high end of the regression of Fig 7, resulting in a larger bivariate fit slope. Decreasing the P3b values by ~2-4 ppb with respect to the DC8 values places this cluster of CO values nearer

the 1 to 1 slope. On the other hand, in the case of the O₃ regression, this period of the 1-2 ppb difference in O₃ measurements occurs approximately mid-range in the O₃ values (~55 ppb) resulting in a much smaller impact on the bivariate fit slope. Figure 10 shows a 200 second period of time encompassing the largest peaks in both Figures 8 and 9. Note that not only do the structures look very similar for the same compound, but a time lag on the order of a few seconds for the DC-8 can be observed in both O₃ and CO measurements. The inclusion of these similar, rapid, and highly precise measurements on both aircraft adds greatly to our confidence that two aircraft intercomparisons can be made highly credible while still requiring only a small amount of additional mission resources. One brief exception discussed in the previous section in which NO and the associated OH varied in coincidence with each other but differently between the 2 aircraft despite similar O₃ and CO values still suggests some degree of caution. As more measurements are added and the speed of existing measurements increases particularly for relatively short lived compounds, chemical differences such as that noted above, which should in general be expected to occur a small fraction of the time unless the measurements are completely collocated, will be more easily identified.

All of the measurements shown in Figures 2-7 introduce such a small potential error into model comparisons between aircraft compared to the uncertainties introduced by other measured quantities that any minor differences can in most cases probably be ignored. It is not clear that this is the case for the next set of measurements. Figure 11 shows a comparison of NO measurements made on the P-3B by *Kondo et al.* [1997] versus those on the DC-8 by *Sandholm and Tan* (this issue). The slope of the bivariate fit through the origin is 0.902 with a similar slope for the unconstrained fit and a near zero

intercept. Thus, there appears to be a systematic difference with the DC-8 instrument measuring about 10% higher on average but still with a significant amount of scatter. The dark dotted lines provide an approximate upper and lower bound for expected data scatter for an average slope of 1 (the 1 to 1 line \pm the square root of the sum of the squared errors for both instruments plus the detection limits) using the errors given in the TRACE-P data table. Assuming that the error limits given in this data table are 2σ error limits (in the case of NO they are $\pm 10/20\%$ for P-3B/DC-8), few (about 5%) of the points should be expected to fall outside of this set of lines. For NO this number is closer to 20%, suggesting that one or both of the error limits may be somewhat underestimated. To better understand measurement differences, Figure 12 shows a plot of the two NO measurements as a function of time during the 3 intercomparison periods. While some of the larger discrepancies are associated with rapid changes in NO, similar discrepancies occur during periods of slow NO changes. Measurement differences are also not consistent: for example, in the first flight the DC-8 measurements sometimes are high and other times low. In the second flight, the P-3 values are consistently higher, and in the third flight, consistently lower. This suggests some type of a shift in calibration or sensitivity between flights, which can not easily be compensated for when comparing models from the 2 aircraft. While an average slope of 0.9 suggests fairly good agreement and potential 10% effects on modeling, trying to compare relative NO concentration between the second and third set of flights can lead to discrepancies of a factor of one and a half, with far more significant effects on model interpretation.

Figure 13 shows a comparison of PAN data measured on the P-3B by Flocke and Weinheimer (this issue) versus those from the DC-8 measured by Singh *et al.* (this issue).

The constrained slope is 1.13, which suggests average agreement within the stated error (-10+5/20% for P-3B/DC-8). The two dotted lines show that individual measurements are outside of the expected error limits about 40% of the time, but only by a small amount. This is largely due to the slope not being quite equal to 1 around which the error lines are centered. The small amount of scatter around the average slope is actually quite impressive since the P-3B measurements have a sample integration time of 1-2 seconds compared to 120 seconds on the DC-8. Figure 14 shows P-3B and DC-8 PAN measurements as a function of time for the second two comparison flights. Here there appear to be no surprises; the slope of about 1.13 describes well the average agreement with no large deviations. The P-3B values are nearly always either above or only slightly below the DC-8 values, and if the dotted error limit bars were centered around the slope of 1.13, essentially all data would fall within the area they bracket. Whatever the cause calibration, interferences, or sampling losses, average differences between measurements appear to be consistent in time, relatively small, and should be much easier to deal with in model comparisons between aircraft.

The final compound in the group is HNO_3 . Figure 15 shows the measurements of HNO_3 on the P-3B by *Zondlo et al.* (this issue) versus those measured on the DC-8 by *Talbot et al.* (this issue). Note that while agreement is fairly good and the average slope is close to 1, the scatter and stated errors ($\pm 25/15\%$ for p-3B/DC-8) are quite large. Again, about 40% of the points fall outside of the dotted error limits, suggesting that one or both of the stated errors are underestimates. Figure 16 shows both the P-3B and DC-8 measurements plotted as a function of time. Note that while measurements appear to track each other, there appears to be somewhat of a bias for higher values being measured

by the DC-8 instrument during the second intercomparison period (3/24), while during the third intercomparison (4/9) the higher values were measured far more frequently on the P-3B. The large difference between the unconstrained bivariate fit and the fit forced through the origin in Figure 15 largely arises from this relative difference between flights, combined with essentially all of the low concentration measurements being made in the last intercomparison period. As in the case of NO calibrations, interferences or losses appear to vary during the mission, resulting in greater difficulty in comparing model results obtained independently for each aircraft.

The next group of measurements which consists only of OH and HO₂/RO₂, show significant discrepancies but both also have relatively large error bars. Figure 17 shows the OH concentration measured on the P-3B by *Mauldin et al.* (this issue) versus that by *Brune et al.* (this issue). The relatively large scatter is consistent with the larger, stated uncertainties. The constrained bivariate fit has a slope of 1.50 which is also quite large, but well within the combined ($\pm 60/40\%$ for P-3B/DC-8) error limits of the two measurements. About 10% of the measured values fall outside of the dotted error lines. These are all on one side, as shown in Figure 17, with no points even close to the other error line. This number is larger, but somewhat consistent with expected scatter, except that it is all biased to one side of the error range because the average slope is 50% above 1. The dark dot-dashed lines in this figure show error limits that are centered around a slope of 1.50, but represent error limits that are only 60% as large as the dotted line centered around a slope of 1. Note that only about 5% of the data points fall outside of these lines. This suggests that the scatter of the data from both instruments is probably better than that suggested by the stated error limits which is consistent with the precision

being better than the absolute accuracy, but that there seems to be a calibration problem associated with these measurements. This is not at all surprising, since OH measurements are inherently difficult to calibrate, due to the lack of stable standards and the rapid reactions of OH on surfaces. In fact the accuracy of OH measurements is largely determined by the uncertainties associated with absolute instrument calibration. It is also interesting to note that despite common concerns about the sensitivity of OH measurements, Figure 17 shows that the data around 1×10^6 molecules cm^{-3} falls between either set of error lines at least as well as does data at higher concentrations. Figure 18 shows P-3B and DC-8 OH concentration as a function of time for the three intercomparison periods. Also shown is a solid line which is proportional to the product of $j\text{O}_3$, O_3 and H_2O (OH production) and the average NO concentrations on both aircraft (interconversion of HO_x). Both of the OH instruments appear to track some of the larger changes in production and NO such as the largest NO peak in the first leg and production increase in the last leg but both also show some inconsistencies. The overall discrepancies are reasonably consistent in time with the P-3B measurements either higher than or equal to those of the DC-8, except for a brief period at the beginning of the second intercomparison period. Many of the largest discrepancies appear to have occurred during flights in the boundary layer, such as in the first half of the first comparison period and the last few points in the last comparison period. These were also periods in which the largest differences and changes occurred in $j\text{O}_3$ values, though the $j\text{O}_3$ changes alone were far too small to explain these differences. Mechanistically, no explanation can be provided for why OH should vary significantly with small actinic flux changes, and what is observed may be purely measurement scatter; however, it may be

desirable to carry out future OH intercomparisons in relatively cloud-free areas if possible, at least until such discrepancies can be better understood.

The relatively large discrepancies between the two measurements require that caution be exercised when comparing model results from the two aircraft. The percent discrepancies shown in Figure 18 have no clear altitude, time, or concentration dependence, and thus Figure 17 provides a reasonably complete review of expected differences.

Measurements of HO₂/RO₂ can not be compared directly, because HO₂ plus RO₂ was measured on the P-3B while HO₂ was being measured on the DC-8. In order to compare these measurements, the arbitrary but reasonable assumption was made that the RO₂ concentration was 80% of the HO₂ concentration throughout the measurement period. Therefore, Figure 19 is a plot of 55.6% of the HO₂ + RO₂ measured on the P-3B by *Cantrell et al.* (this issue) versus HO₂ measured on the DC-8 by *Brune et al.* (this issue). A slope of 1.35 is seen to fit the average data, but it should be remembered that the value of this slope is a relative number that depends on the ratio of RO₂ to HO₂ which presumably is not even a constant throughout the intercomparison period. About 10% of the data points fall outside of the dotted error lines and this can be reduced by a factor of 2 if the error lines are centered symmetrically around the slope of 1.35 so the scatter is approximately consistent with the stated uncertainties ($\pm 35/40\%$ for P_3B/DC-8). Figure 20 shows both the P-3B HO₂ plus RO₂ measurements and the DC-8 HO₂ measurements plotted as a function of time during the first and last intercomparison periods. The sum of HO₂ and RO₂ is plotted in this figure because the RO₂ percentage may vary with altitude, which is shown by the dashed line. The P-3B HO₂ + RO₂ is significantly higher

than the DC-8 HO₂ throughout the first comparison and the first half of the third comparison and then became approximately equal to the DC-8 HO₂ for the last half of the comparison. Since both the beginning of the first comparison and the end of the last comparison were in the boundary layer, there appears to be no simple altitude trend. The relatively large stated uncertainties associated with these data and the associated scatter shown in Figure 19 combined with the fact that the amount of HO₂ in the HO₂ + RO₂ measurement is unknown make this comparison particularly difficult. Model comparisons can be made directly to either HO₂ or HO₂ plus RO₂, thus there is no inherent problem associated with measurement/model comparison. Caution should again be exercised in comparing results from the two different aircraft.

The final group consists of three P-3B and DC-8 measurements that were compared include NO₂, SO₂ and PPN. These all had slopes that differed from 1 by a factor of 2.5 to 3.5, with most of the data for SO₂ and PPN outside of the expected error range. Figure 21 shows the NO₂ concentration measured on the P-3B by *Kondo et al.* [1997] versus the DC-8 value measured by *Sandholm and Tan* (this issue). While most of the NO₂ points do fall within the error limits (± 32 -50/40% for P-3B/DC-8) shown in Figure 21, the slope of the unconstrained bivariate fit is approximately zero. It should also be noted that most of the data in this figure is below the stated detection limit for the P-3B instrument (13 pptv), and thus should not be intercompared, except that DC-8 data from the same time period suggest that the NO₂ was 2-5 times the P-3B detection limit. Also, the comparison was dramatically worse during the second intercomparison period than in the first (no comparison data for the third), suggesting a high degree of inconsistency in either measurement or calibration on the part of one or both instruments.

Comparing NO₂ results of the first comparison period to those of NO shows reasonable agreement for both NO₂ measurements, while in the second the Kondo NO₂ measurements appear to better track the gradual NO decline with time.. Clearly, more work is needed to resolve large differences in the measurement of this important compound, some type of adjustment needs to be made to NO₂ model values when comparing results from the two aircraft, and much could be learned about detection limits versus calibration problems if future comparisons contain NO₂ concentrations well above the NO₂ detection limit.

Figure 22 shows the SO₂ concentrations measured on the P-3B by *Thornton et al.* [2002] versus those measured on the DC-8 by *Talbot et al.* (this issue). The P-3B values are higher than the DC-8 values by up to an order of magnitude, except for one brief period in the middle of the last intercomparison period. These discrepancies are far beyond the error limits ($\pm 2-3/20\%$ for P-3B/DC-8) or detection limits of either instrument and need to be investigated further.

Figure 23 shows PPN measurements made on the P-3B by *Flocke and Weinheimer* (this issue) versus the DC-8 measurements made by *Singh et al.* (this issue). Again, agreement is poor and nearly all points are well beyond the error lines ($-10+5/30\%$ for P-3B/DC-8). It should also be noted, however, that nearly all measurements are within a factor of 3 of the detection limits (5/1pptv for P-3B/DC-8) for both instruments. Since PPN is measured by the same instruments used to measure PAN, but is observed to be at so much lower concentration, the influence of this discrepancy on overall model predictions is probably small. While efforts should be made to better understand and remove these discrepancies, at least an equal amount of effort needs to go

into measuring and intercomparing measurements for other PAN-like compounds, for which there are even less data.

DC-8 Intercomparison Results

The intercomparisons discussed in this section extended throughout the TRACE-P mission and included measurements of quite clean and also highly polluted air masses. Thus, the number of comparison points and also the range of these measurements tend to be much larger than those in the previous section. Figure 24a shows formaldehyde concentrations measured by *Heikes et al.* (this issue) versus those measured by *Fried et al.* (this issue). Figure 24a shows the whole data set, while Figure 24b shows just formaldehyde values below 600 pptv so that the majority of the data can be seen more clearly. From both figures, it is clear that many of the data points fall outside of the $\pm 21\%$ error bars centered around a slope of 1. If similar error bars are centered around the average slope of 1.49 (dotted/dashed lines) which is strongly driven by the highest observed concentrations, more of the very highest concentrations fall within the bracketed region, but little improvement is observed for data in the 2-3 ppbv range and below. An additional concern is also seen in Figure 24b. There are a large number of points that are at the stated detection limit for one instrument, such as those plotted for the Heikes instrument at 50 pptv or the Fried instrument below 58-80 pptv, which have a companion measurement by the other instrument which is far above the detection limit ($\pm 15/12-15\%$ for Heikes/Fried). This can be seen in Figure 25 which allows discrepancies to be more clearly seen near the detection limit. This figure plots all the CH₂O comparison data acquired by the two instruments on the DC-8 as differences (Heikes –

Fried) versus the average of the two. The total combined uncertainty limits (2σ) are shown by the solid black lines, and these were calculated from the quadrature addition of the total uncertainties from the two instruments (Heikes: 15% of Concentration + 50 pptv; Fried: $[(\text{LOD})^2 + (\text{Systematic Term})^2]^{1/2}$). This figure shows three different regions. The first region indicates that 61% of the comparison points yield differences within the combined uncertainty limits. The upper region contains 26% of the measurements and the low region has 13%. These results suggest several areas that need to be addressed. There appears to be an overall inlet/instrument calibration problem at high concentrations, but at low concentrations correlated data scatter is larger so that nearly half of the data points shown in Figure 24a and b fall outside of the expected error limits. This suggests a measurement problem well beyond the stated error limits and/or detection limits for one or both instruments. In an attempt to provide some additional insight into how the concentrations of formaldehyde and other compounds in this section changed in various types of air masses, CO will be plotted as a function of time along with the other measurements made in this section.

The compounds being compared in this section (formaldehyde, acetone, methylethylketone, acetaldehyde, and methanol) are all products of some type of hydrocarbon oxidation. In some cases they may also have direct emission sources but even these would be expected to be associated with urban/industrial or biomass burning plumes which typically also contain elevated CO. CO has a relatively long atmospheric lifetime with much of its decline in concentration with time in plumes due to dilution rather than destruction. The atmospheric lifetimes of the compounds being compared in this section are varied but are generally much shorter than that for CO with chemical

production and loss mechanisms also quite different. Thus, a very high degree of correlation with CO is not expected. On the other hand as a tracer for plumes containing a large amount of reactive carbon some significant degree of correlation with CO would be expected and a total lack of correlation would seem difficult to explain. The following correlation plots are not being presented as a quantitative test of instrument performance but rather as a fairly general qualitative means of assessing measurement discrepancies for cases where correlation plots for the same compounds differ significantly. When looking at the correlation plots for the next five compounds it should be noted that where agreement between similar measurements improves this seems to be reflected in a better correlation with CO for these measurements. This appears to be the case both for comparing one compound to another or in some cases, the agreement and correlation with CO when comparing low and high concentration of a single compound. For example, higher formaldehyde appears to correlate better both between instruments and with CO for both instruments. Also, acetone and methylethylketone appear to show better instrument to instrument agreement and the Apel/Riemer data appears to show better correlation with CO for these two compounds than for the other two compounds. The Singh data shows reasonable CO correlation for all compounds.

Figures 26a and b show two correlation plots of formaldehyde with CO. The first shows the data from Heikes (a) and the second that from Fried (b). The correlation appears to get better at higher concentrations for both data sets and is probably somewhat less scattered for the Fried data particularly when CO is below 200-300 ppbv. The latter becomes more obvious if the correlation data are expanded so that all points are observable in the areas that are at present saturated with points in Figure 26a and b.

Correlations are still not good, however, and as suggested by direct correlation between similar measurements low concentrations again pose the greatest problem. This is major concern because much of the mission data is either below or within a factor of 2 or 3 of the limit of detection of these instruments. Therefore, additional sensitivity is badly needed for measurements in relatively clean air masses. There are also some discrepancies that need to be addressed throughout the concentration range, but many of these may be difficult because they are sporadic in nature. There are no multi-aircraft model comparison issues that need to be addressed for any of the measurements in this section. From a practical modeling standpoint, it should be noted that the Fried formaldehyde data coverage is about 53 % (34% above LOD), while the Heikes data is available for about 26% (16% above LOD) of the time.

The 4 compounds in the final set were all measured independently by *Apel et al.* (this issue) and *Singh et al.* (this issue). These data are particularly difficult to compare because the Apel/Riemer data is typically measured with 10 to 60 second integration times for low and high altitudes respectively, while the Singh measurements are 180 second integrations. This is discussed in more detail by Apel et. al, (this issue). Figures 27-30 show the data of Apel and Riemer versus Singh for the compounds acetone, methylethylketone, acetaldehyde, and methanol respectively. The slope of 1.33 shown for acetone is somewhat beyond that expected from the combined uncertainties ($\pm 3/20\%$ for Apel/Singh), but not surprising considering the integration time differences. A contributor to this difference is the disparity observed in the calibration standards as analyzed by Riemer et al. [this issue] during the study. The Apel-Riemer standard yielded values that would result in 12.5 % higher values for acetone than if the Singh standard

values were used. In the other three cases, the agreement cannot be said to be good in that the slopes of the bivariate fits that are forced through the origin range from about 1.6 to 2.1, and the slope of the unconstrained fits are even larger. These slopes are much larger than those expected from the combined uncertainties of both instruments ($\pm 3/20\%$ for methylethylketone and $\pm 7/25\%$ for acetaldehyde and methanol for Apel/Singh) also, most of the data fall outside of the dotted error lines, and even if these lines are centered around the average slopes, no dramatic improvement is observed. In some cases, however, such as for acetone and methylethylketone there does appear to be some reasonable correlation. Unlike formaldehyde, though, agreement does not seem to improve very much with concentration. Thus, lack of sensitivity does not appear to be a significant contributor to the observed discrepancies. Also, a constant calibration error does not in general appear to be the major problem. Rather, discrepancies are highly variable over the entire measured range, with all 4 of the Apel/Riemer measurements higher than those of Singh most of the time. This could result either from an interference that could sporadically enhance signal, or a variable sampling loss, which could either reduce the concentration of the compound being measured or possibly delay the instrument response time. A post-mission instrument evaluation [Apel et al., this issue] revealed interferences in the acetone and acetaldehyde measurements of Apel/Riemer but these have already been corrected for the comparison data shown. Any of the former problems could dramatically degrade the data correlation between these two instruments. The same problems, however, would be expected to degrade correlations with other related compounds as well. Figure 26(c-j) shows the concentration of each of the above compounds plotted against that of CO. While some compounds show a better correlation

with CO than others, it is reasonable to assume that all should have some degree of positive correlation even if CO is only assumed to be a tracer for Asian plumes. In several cases, there appear to be significant differences in the degree of correlation with CO. Some lack of correlation may be unavoidable in the Apel-Riemer data because of the 10 – 60 second time base used. Their sampling time base is dependent on altitude. Most plumes are observed at relatively low altitude; this is where the time base for Apel/Riemer is the shortest, often less than 20 seconds. Relatively high variability is observed for CO over the 1 minute time periods [Riemer et al, this issue], perhaps precluding excessively tight correlations even in plumes. Better correlation with CO is expected for shorter-lived compounds that have no significant non-combustion sources. For longer-lived compounds such as acetone and methanol that have significant additional sources, one might expect a poorer correlation, particularly outside of plumes. This issue is also addressed by Riemer et al., [this issue]. It should be noted that for all 4 of these compounds and also for formaldehyde, the plots that appear to show the best correlation come from instruments that on average measured lower concentrations. If there were compounds that were highly correlated with plumes and CO which caused interferences in these instruments, they could enhance the observed correlation with CO, but they would also presumably lead to higher not lower measured concentrations. Thus it seems unlikely that the better correlations observed are a result of measurement interferences.

Conclusions

The primary conclusion of this paper is that multi-aircraft intercomparisons can be made credible and therefore extremely valuable. If such comparisons are included as

part of a field campaign, they can be accomplished with little additional effort and can directly provide comparison information on the exact instrument configuration used on the mission and its response to conditions encountered during that mission. The availability on multiple aircraft of rapid, high precision measurements of compounds such as CO and O₃, which can vary over a relatively large dynamic range, contributed much to the credibility of this intercomparison. Good agreement of these measurements by itself does not assure that identical air masses are being sampled by both aircraft as discussed for complementary variations in NO and OH which differed between aircraft. This rare incident, however, occurred in the boundary layer during the first intercomparison leg for which the average aircraft separation was still a few km. In later flights, when the aircraft separation was reduced to well below 1 km, the NO agreement was always much better than the factor of two observed briefly during the first comparison 3/4/2001 at around 1:24. Discrepancies observed in NO at the 20% level were present when flying through both structured air masses and what appeared to be relatively uniform air masses. If the precision and calibration of these NO instruments could be improved so that agreement between aircraft (not necessarily absolute accuracy or measurements near the detection limit) was consistently about 5%, the shorter lifetime and much larger dynamic range of NO would provide an additional major improvement in assessing air mass similarity. Fast water measurements on both aircraft would also provide similar but complementary insight into air mass similarity. At present, without very good agreement in NO or water measurements, it is not clear that an event of relatively short duration (small distance) would be detectable if it only involved differences on the order 20% of in relatively short atmospheric lifetime species.

Averaged over a significant portion of an intercomparison flight (for example 100-200km), persistent differences in chemical concentrations seem unlikely particularly with an aircraft spacing of only a few tenths of a km, however, at the end of the third intercomparison O₃ and CO measurements showed such differences. While the TRACE-P comparison provides much new general insight into measurement differences and future needs, there still remains concerns about just how similar the air masses were that both aircraft sampled from. In the future, it is hoped that the ever expanding development of more rapid measurement capabilities (including water) combined with experience from previous comparisons and more long range planning of intercomparisons (allowing consistent aircraft separations of only a few tenths of a km or less) as an integral part of field campaigns will make future intercomparisons even more informative.

The results of this comparison were quite varied. The first group of measurements agreed so well that additional improvements would advance the mission science objectives little except where even faster measurements are needed, such as for flux studies. The second and third groups of measurements including NO, PAN, HNO₃, OH and HO₂ showed good promise, particularly PAN, but improvements in all of these would significantly advance scientific goals. OH and HO₂ discrepancies make it particularly difficult to intercompare mission results. In the case of OH, absolute instrument calibration would appear to be an area for improvement. Similar concerns exist for HO₂ and RO₂, but the situation is more clouded by the inability to directly compare results. Hopefully in the future these two instruments can be compared when both are measuring HO₂ or HO₂ + RO₂ or both. The final group of two-aircraft comparisons suggested that at least one of the instruments measuring NO₂, SO₂, and PPN

was either too close to its detection limit or in error. These large discrepancies need to be resolved if these instruments are going to contribute to future joint aircraft measurement and modeling efforts. Additional insight into ongoing concerns about NO₂ discrepancies and model comparison are discussed in Olson et. al. (2001) and Kondo et. al (this issue). While the comparisons of instruments that were solely on the DC-8 during TRACE-P generally appeared to show somewhat poorer results, they were also subjected to a far greater diversity of air masses. Two aircraft comparisons in plumes would be highly desirable in the future, but will require far more planning and some luck. One of the major areas of improvement needed for the instruments that were only on the DC-8 is a more sensitive measurement of formaldehyde. Data coverage is significantly limited by measurements at or below the limit of detection, and far more measurements are within a factor of 2 or 3 of this limit. Additionally, the potential for interferences, inlet effects, and possible calibration problems in the oxygenated hydrocarbon measurements needs to be more fully explored.

As stated previously, the results of this intercomparison should be viewed as a starting point for achieving a better understanding of instrument operation and aircraft measurement problems. This text is specifically not intended to provide a critical review of individual instrument operation, but rather to summarize where additional effort is needed and as a brief guide to modelers who are using data from both aircraft or from the DC-8 where multiple measurements are available. This is the first time that most of these instruments have been compared on an aircraft, and for several instruments only the first or second time that they have flown. There was somewhat of a tendency for the largest discrepancies to be associated with measurements involving at least one new instrument

or measurement technique. This is not meant to suggest that the newer techniques are in error, but rather that very different measurement techniques are more likely to disagree than are similar techniques being used by two different investigators. It is also more likely that agreement will be observed for two instruments that have been compared before than for one or more new instruments which have never been compared. It is when good agreement is achieved between two or more dramatically different measurement techniques using independent calibration methods, however, that the most credible measurement validation is provided. Thus, the development and intercomparison of unique new measurement techniques needs to be encouraged.

Acknowledgements

The authors wish to thank the National Atmospheric and Space Administration (NASA) for their support of the many measurements discussed in this paper through the Global Tropospheric Experiment (GTE) program.

References

Apel et al. (this issue)

Avery et al. (this issue)

Bartlett, K. B., G. W. Sachse, T. Slate, C. Harward, and D. R. Blake, Large-scale distribution of CH₄ in the western Pacific: Sources and transport from the Asian continent (this issue)

Beck, S. M., et al., Operational overview of NASA GTE/CITE 1 airborne instrument intercomparisons: Carbon monoxide, nitric oxide, and hydroxyl instrumentation, *J. Geophys. Res.*, 92, 1977-1985, 1987.

Brune et al. (this issue)

Cantrell et al., Peroxy radical behavior during TRACE-P as measured aboard the NASA P-3B aircraft (this issue)

Flocke and Weinheimer (this issue)

Fried et al. (this issue)

Heikes et al. (this issue)

Hoell, Jr., J. M., D. L. Albritton, G. L. Gregory, R. J. McNeal, S. M. Beck, R. J. Bendura and J. W. Drewry, Operational overview of NASA GTE/CITE 2 airborne instrument intercomparisons: Nitrogen dioxide, nitric acid, and peroxyacetyl nitrate, *J. Geophys. Res.*, 95, 10,047-10,054, 1990.

Hoell, Jr., J.M., et al., Operational overview of the NASA GTE/CITE 3 airborne instrument intercomparisons for sulfur dioxide, hydrogen sulfide, carbonyl

- sulfide, dimethyl sulfide, and carbon disulfide, *J. Geophys. Res.*, 98, 23,291-23,304, 1993.
- Kondo, Y., et al., The performance of an aircraft instrument for the measurement of NO_y, *J. Geophys. Res.*, 102, 28,663-28,671, 1997.
- Kondo, Y. et al., (this issue)
- Mauldin et al., Highlights of OH, H₂SO₄, and MSA measurements made aboard the NASA P-3B during TRACE-P (this issue)
- Moore et al. (this issue)
- Olson, J. R., et al., Seasonal differences in the photochemistry of the South Pacific: A comparison of observations and model results from PEM-Tropics A and B, *J. Geophys. Res.*, 106, 32,749-32,766, 2001.
- Riemer et al. (this issue)
- Sandholm and Tan (this issue)
- Singh et al. (this issue)
- Talbot et al., Reactive nitrogen in Asian continental outflow over the western Pacific: Results from the NASA TRACE-P airborne mission (this issue)
- Thornton, D. C., A. R. Bandy, F. H. Tu, B. W. Blomquist, G. M. Mitchell, W. Nadler and D. H. Lenschow, Fast airborne sulfur dioxide measurements by atmospheric pressure ionization mass spectrometry (APIMS), submitted to *J. Geophys. Res.*, 2002.
- Vay et al., The influence of regional-scale anthropogenic activity on CO₂ distributions over the western North Pacific, (this issue)
- Weber et al. (this issue)

Zondlo, M. A., R. L. Mauldin, E. Kosciuch, C. A. Cantrell and F. L. Eisele, Development and characterization of an airborne-based instrument used to measure nitric acid during the NASA TRACE-P field experiment (this issue)

Figure Captions:

Figure 1. Plot of altitude for the P-3B and DC-8 aircraft as a function of time throughout each of the 3 intercomparison periods. The middle flight was at a fixed altitude so, the altitude scale was expanded to better show typical height differences. Also shown are approximate horizontal separation between aircraft.

Figure 2. Correlation plot of $j\text{O}_3$ on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = -6×10^{-7}) and solid line is a similar fit forced through the origin.

Figure 3. Correlation plot of $j\text{NO}_2$ on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = -3.4×10^{-4}) and solid line is a similar fit forced through the origin.

Figure 4. Correlation plot of CH_4 on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = -60) and solid line is a similar fit forced through the origin.

Figure 5. Correlation plot of CO_2 on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 9.1) and solid line is a similar fit forced through the origin.

Figure 6. Correlation plot of O_3 on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 0.14) and solid line is a similar fit forced through the origin.

Figure 7. Correlation plot of CO on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = -12) and solid line is a similar fit forced through the origin.

Figure 8. Plot of one second O_3 measurements on the P-3B and the DC-8 vs time during the last set of intercomparison flights. Note that the structure that both aircraft encounter is extremely similar except before the start of the parallel (intercomparison) flights which began about 6:04.

Figure 9. Plot of one second CO measurements on the P-3B and the DC-8 vs time during the last set of intercomparison flights. Note that again the structure that both aircraft encounter is extremely similar.

Figure 10. Plot of one second O_3 and CO measurements on the P-3B and the DC-8 vs time for 200 seconds around the highest peaks seen in figures 8 and 9. Note that while the structure observed by both aircraft is extremely similar for the same compound there is a

time shift of a few seconds. This time shift appears reasonably similar for both O_3 and CO even though the structure observed in these two compounds is quite different.

Figure 11. Correlation plot of NO on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 0.49) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 12. Plot of NO measurements on the P-3B and the DC-8 vs time during the 3 sets of intercomparison flights.

Figure 13. Correlation plot of PAN on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 6.0) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 14. Plot of PAN measurements on the P-3B and the DC-8 vs time during the last 2 sets of intercomparison flights.

Figure 15. Correlation plot of HNO_3 on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 22) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 16. Plot of HNO_3 measurements on the P-3B and the DC-8 vs time during the last 2 sets of intercomparison flights.

Figure 17. Correlation plot of OH on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = -3.1×10^5) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1. The heavy dot/dashed lines show expected error bounds only 60% as large as the heavy dotted lines but centered around a slope of 1.50.

Figure 18. Plot of OH measurements on the P-3B and the DC-8 vs time during the 3 sets of intercomparison flights. The solid line in each intercomparison leg is a crude relative measure of OH production ($jO_3 \times O_3 \times H_2O$ averaged for both aircraft) and the dotted line is the averaged NO.

Figure 19. Correlation plot of HO_2 estimated from the HO_2+RO_2 measurements on the P-3B vs the HO_2 measured on the DC-8. The dashed line is a bivariate fit to the data (intercept = -8.4×10^7) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 20. Plot of HO_2+RO_2 measurements on the P-3B and HO_2 measurements on the DC-8 vs time during the first and last sets of intercomparison flights.

Figure 21. Correlation plot of NO₂ on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 14) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 22. Correlation plot of SO₂ on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 29) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 23. Correlation plot of PPN on the P-3B vs that on the DC-8. The dashed line is a bivariate fit to the data (intercept = 8.1) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 24 Correlation plot of CH₂O measured by Heikes vs that measured by Fried both on the DC-8. The dashed line is a bivariate fit to the data (intercept = -31) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1. The dot/dashed lines are centered around the average slope of 1.49 but with the same range of expected error limits as the heavy dotted lines. Figure (a) shows the whole data set and (b) shows the same data and lines seen in figure 24a but expanded between the origin and 600pptv.

Figure 25 Shows a plot of the differences in Heikes – Fried CH₂O data versus the average of the two. The total combined uncertainty limits (2σ) are shown by the solid black lines, which were calculated from the quadrature addition of the total uncertainties from the two instruments (Heikes: 15% of Concentration + 50 pptv; Fried: $[(\text{LOD})^2 + (\text{Systematic Term})^2]^{1/2}$).

Figure 26. Individual correlation plots of all of the measurements intercompared on the DC-8 vs CO. The same scale is used to compare similar measurements but in some figures up to three data points are not shown in order to expand these scales as much as possible.

Figure 27. Correlation plot of acetone measured by Apel vs that measured by Singh both on the DC-8. The dashed line is a bivariate fit to the data (intercept = -500) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 28. Correlation plot of methylethylketone measured by Apel vs that measured by Singh both on the DC-8. The dashed line is a bivariate fit to the data (intercept = -27) and solid line is a similar fit forced through the origin. The heavy dotted lines show the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 29. Correlation plot of acetaldehyde measured by Apel vs that measured by Singh both on the DC-8. The dashed line is a bivariate fit to the data (intercept= -295) and solid line is a similar fit forced through the origin. The heavy dotted lines shows the approximate expected bounds of data scatter assuming an average slope of 1.

Figure 30. Correlation plot of methanol measured by Apel vs that measured by Singh both on the DC-8. The dashed line is a bivariate fit to the data (intercept= -2579) and solid line is a similar fit forced through the origin. The heavy dotted lines shows the approximate expected bounds of data scatter assuming an average slope of 1.

Fig 1

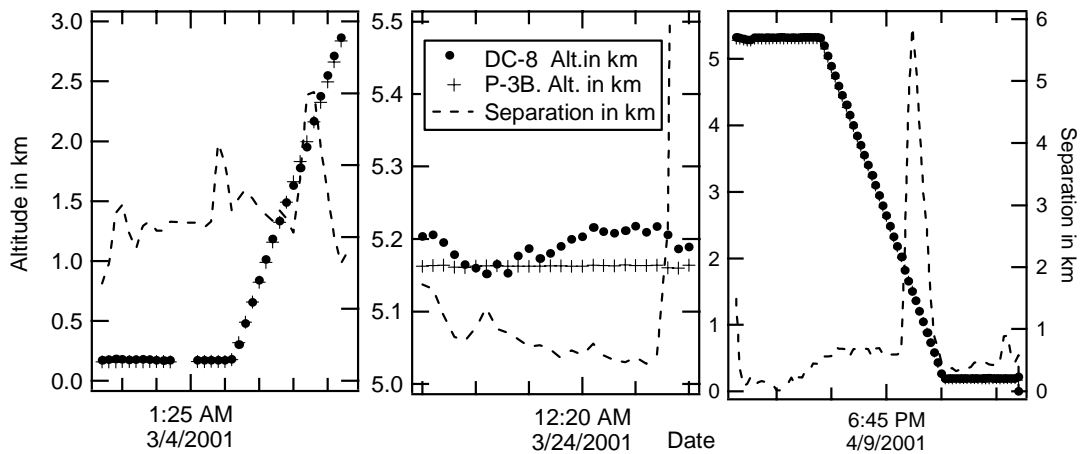


Fig 2

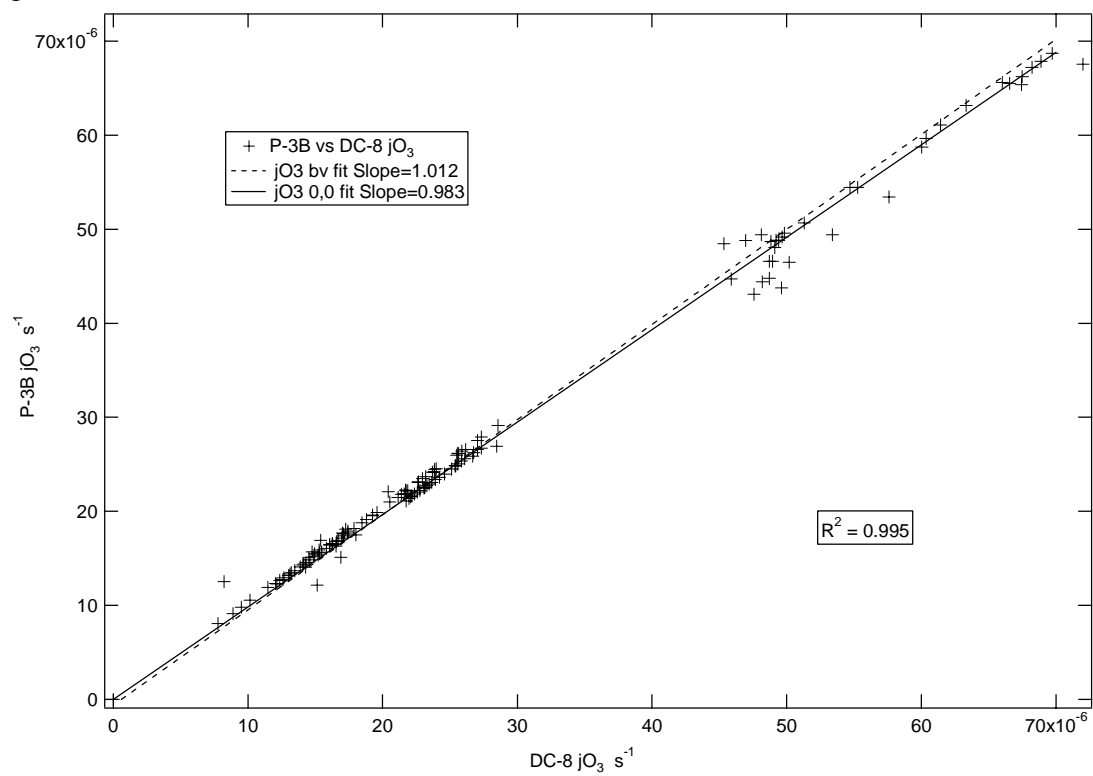


Fig 3

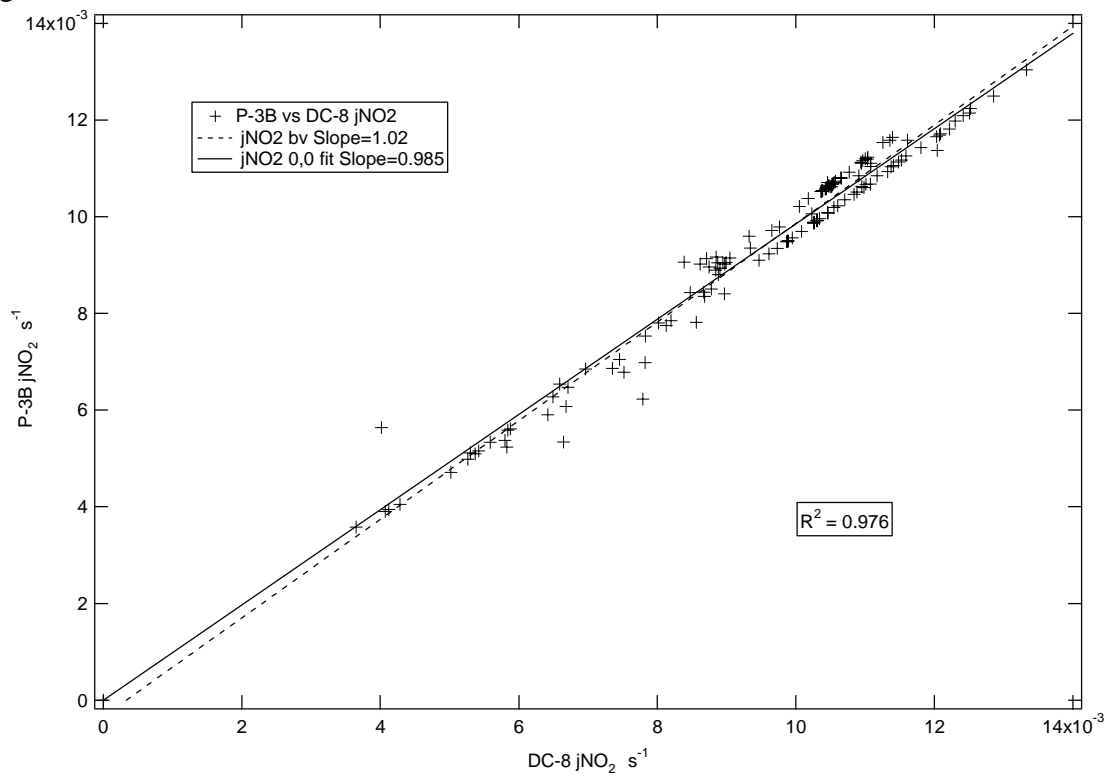


Fig 4

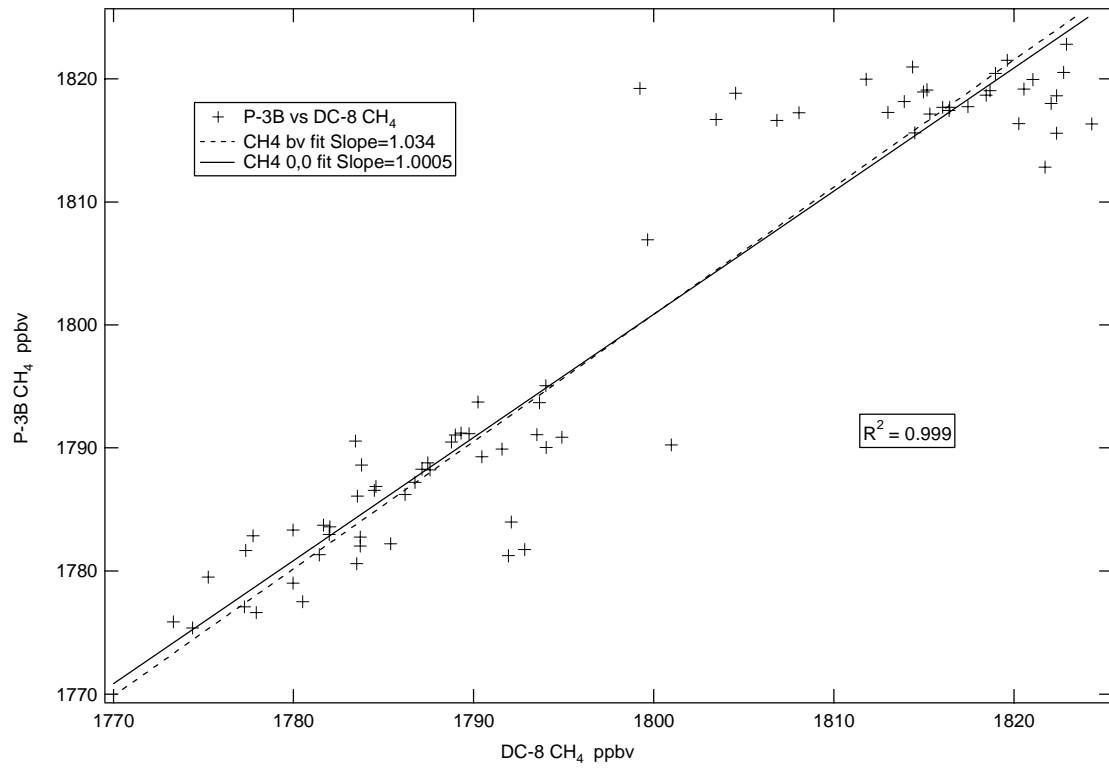


Fig 5

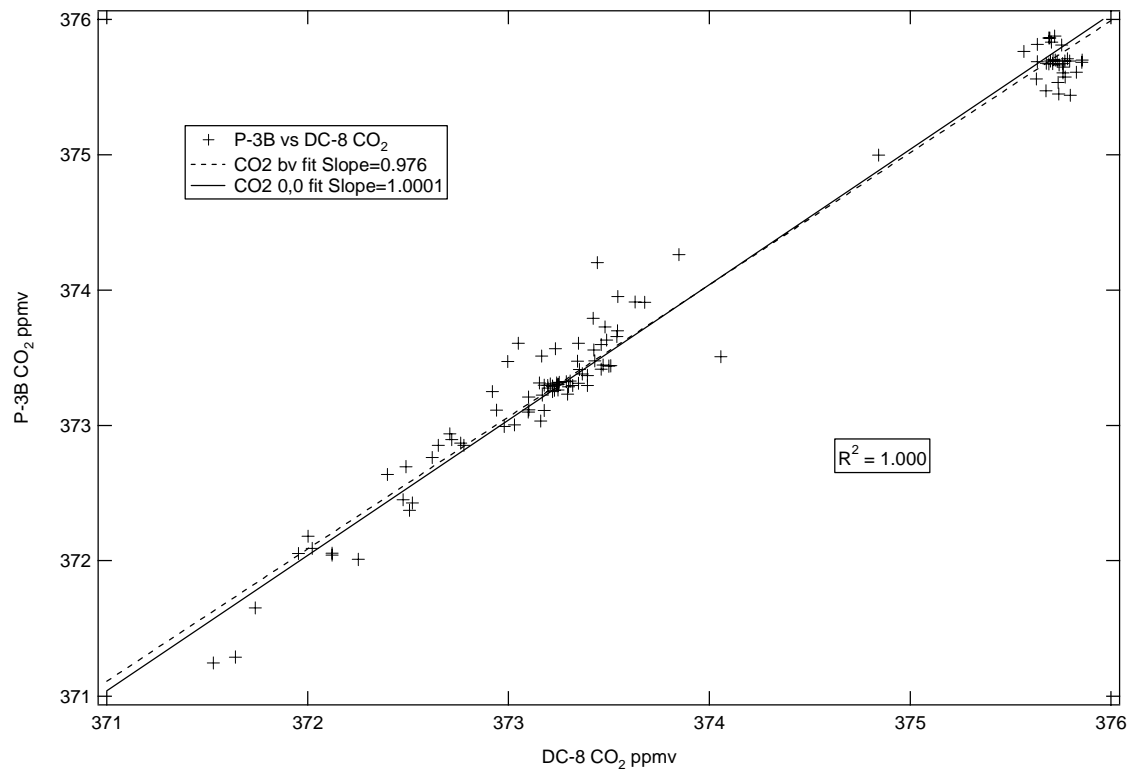


Fig 6

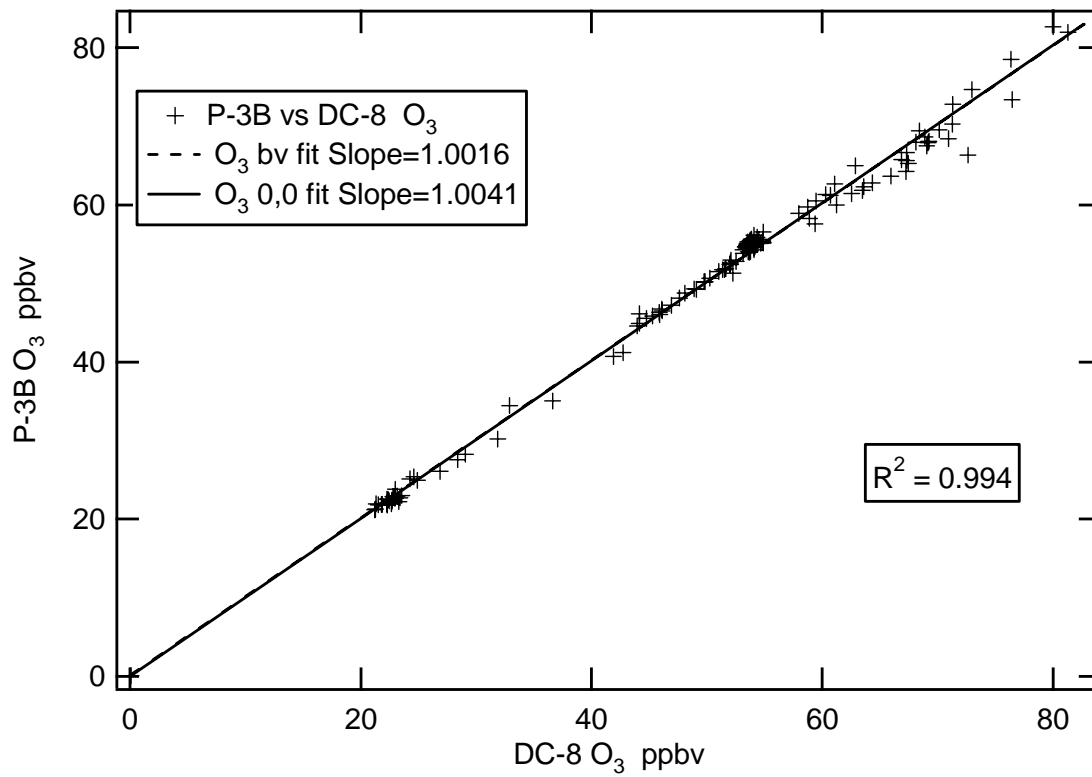


Fig 7

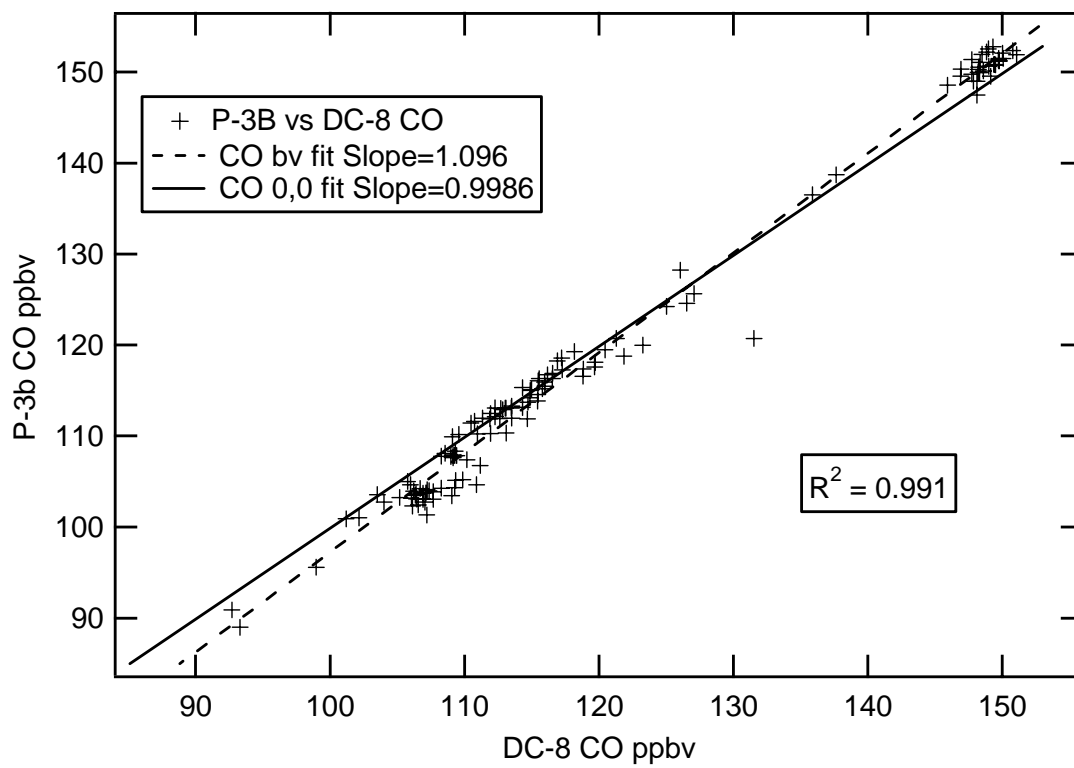


Fig 8

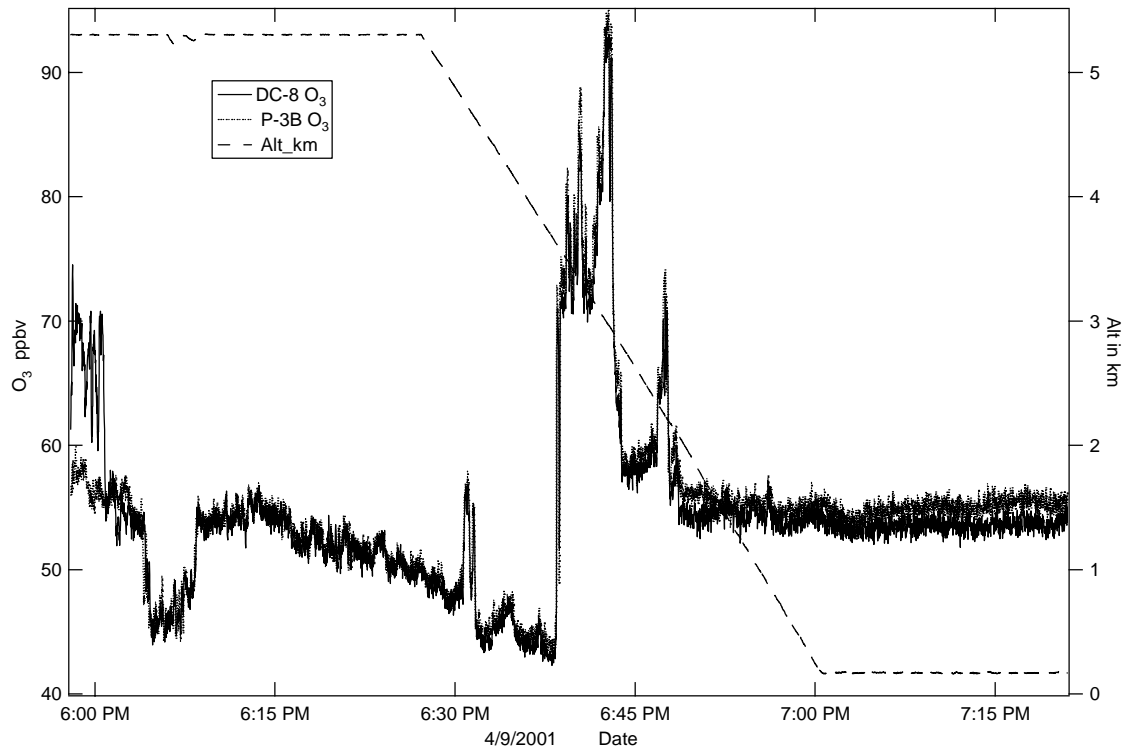


Fig 9

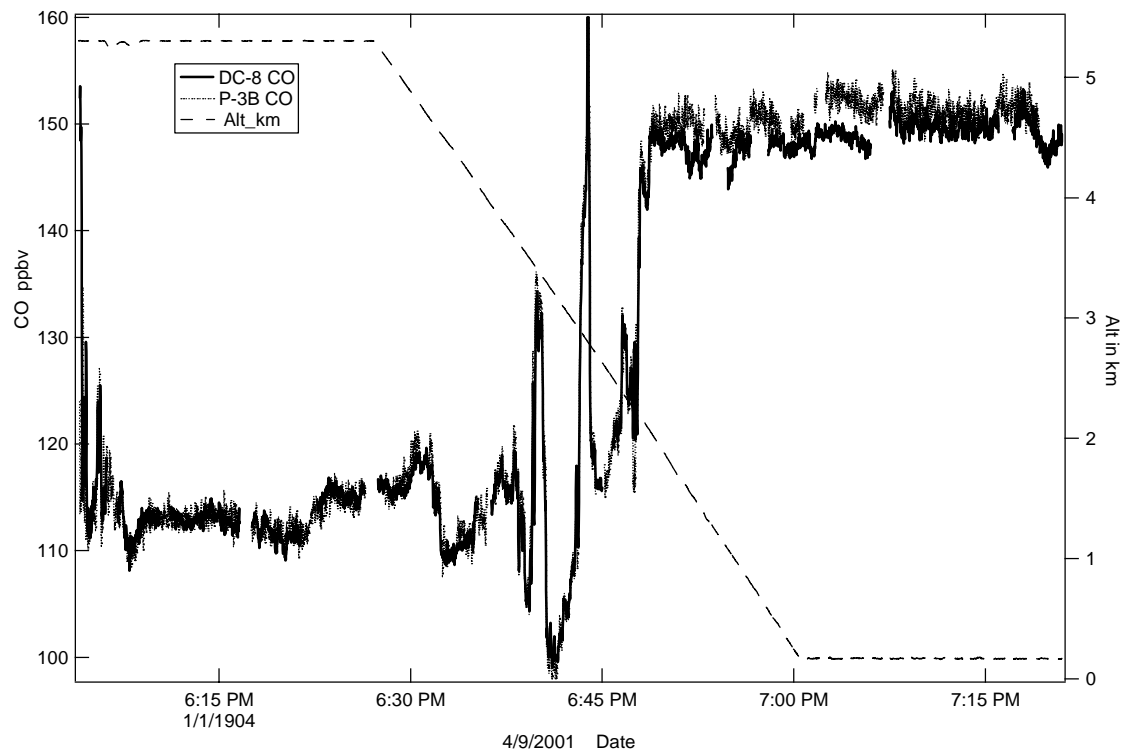


Fig 10

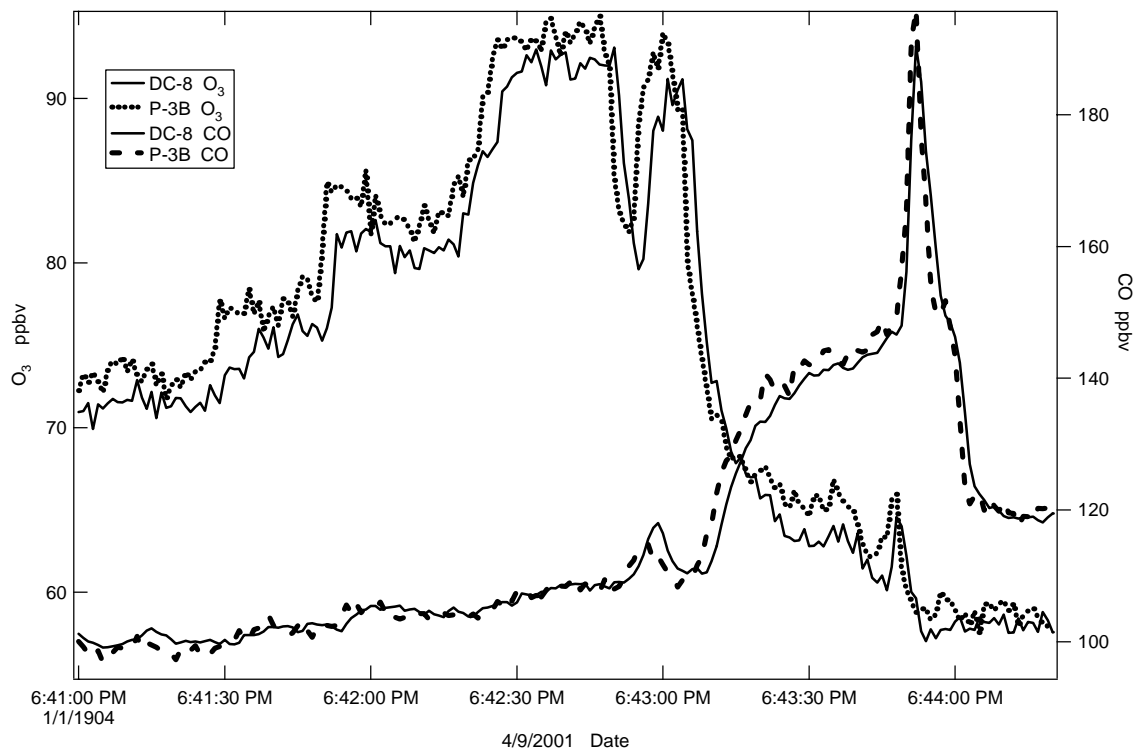


Fig 11

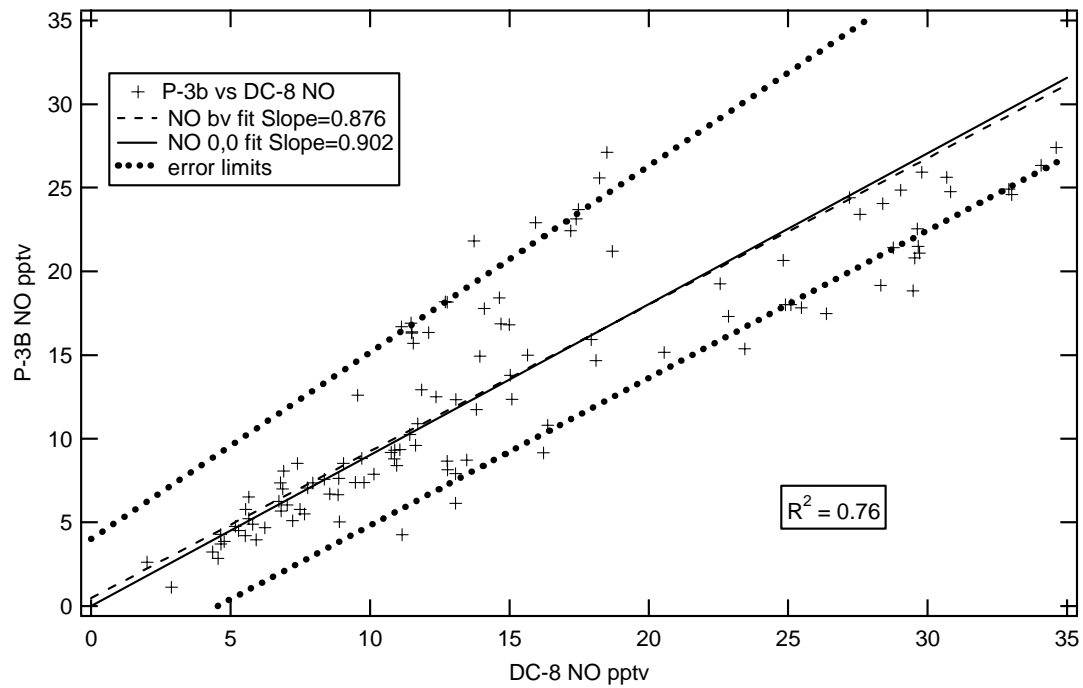


Fig 12

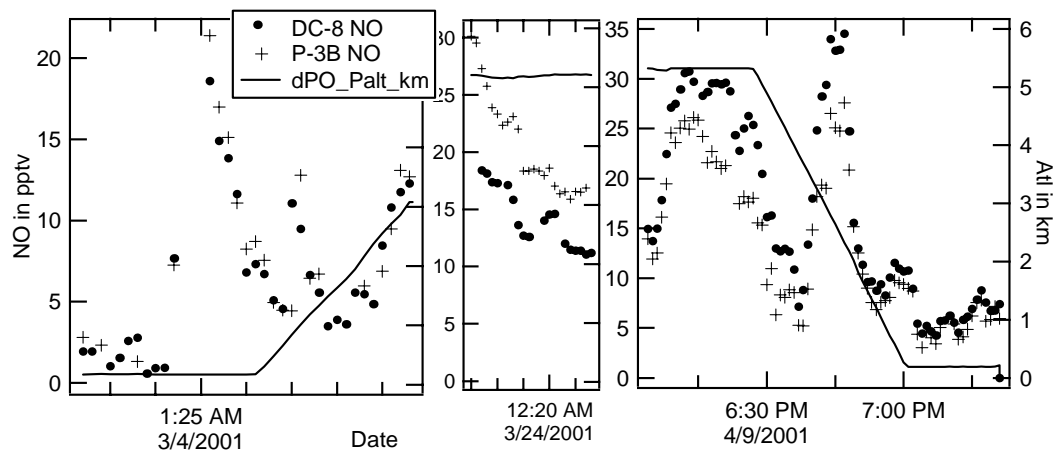


Fig 13

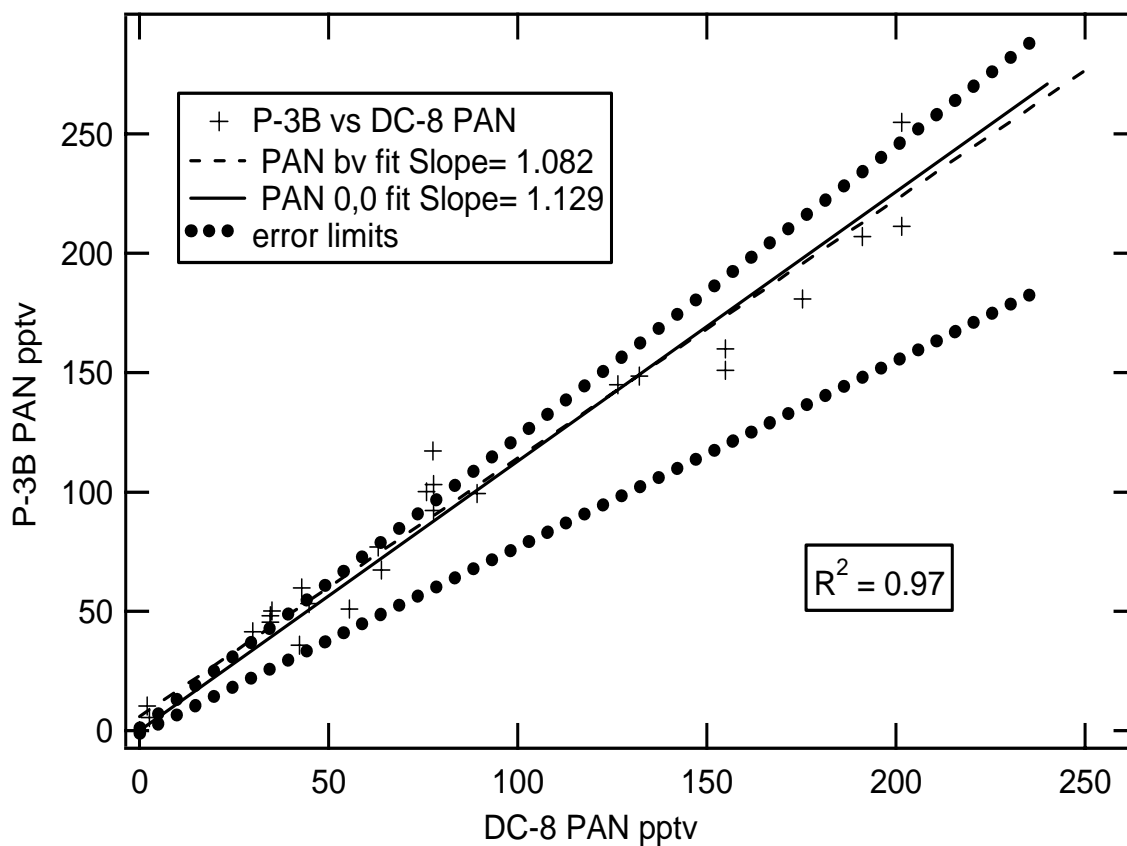


Fig 14

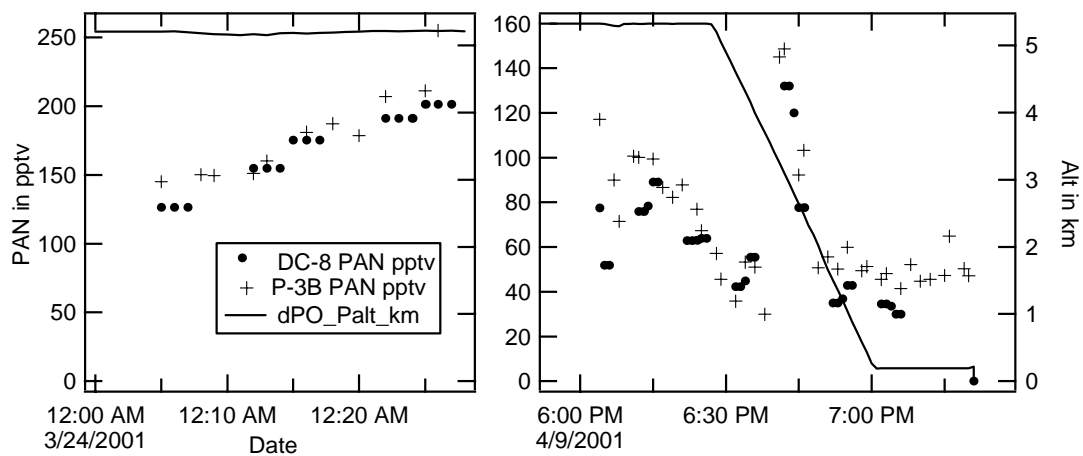


Fig 15

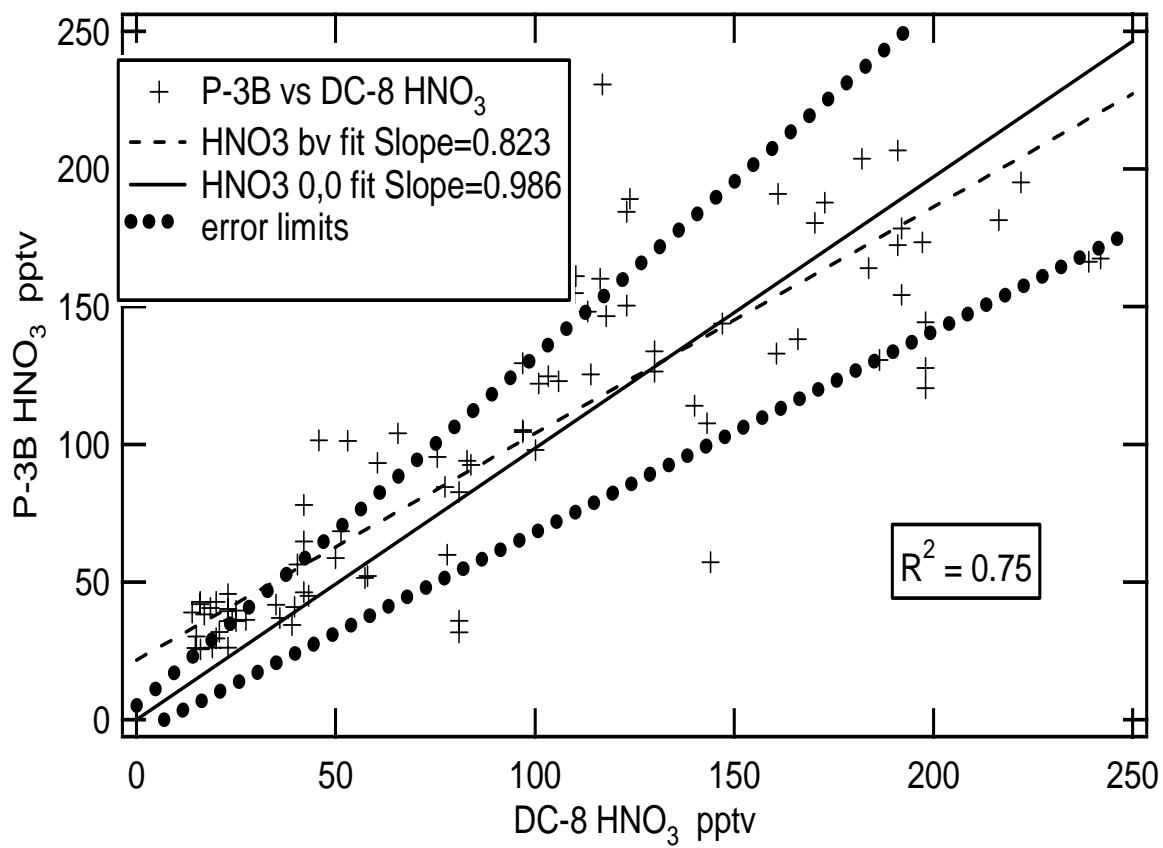


Fig 16

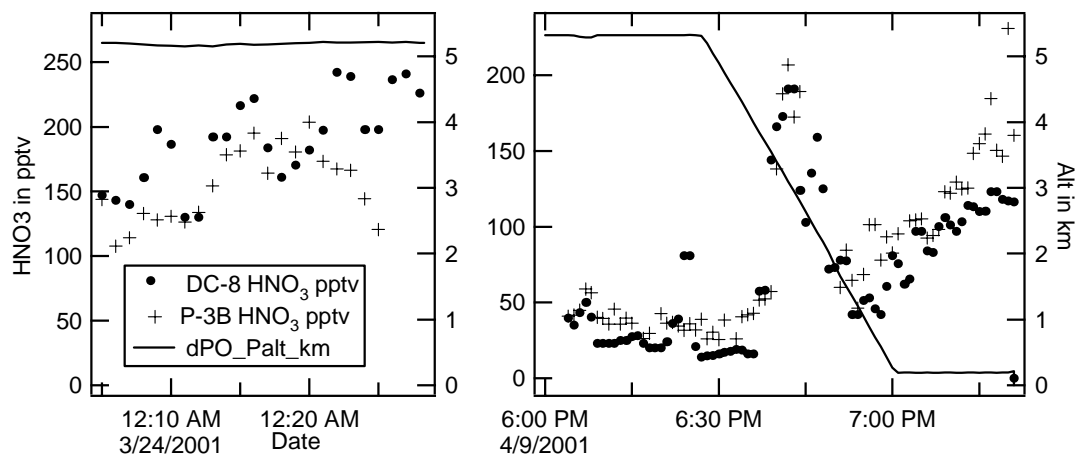


Fig 17

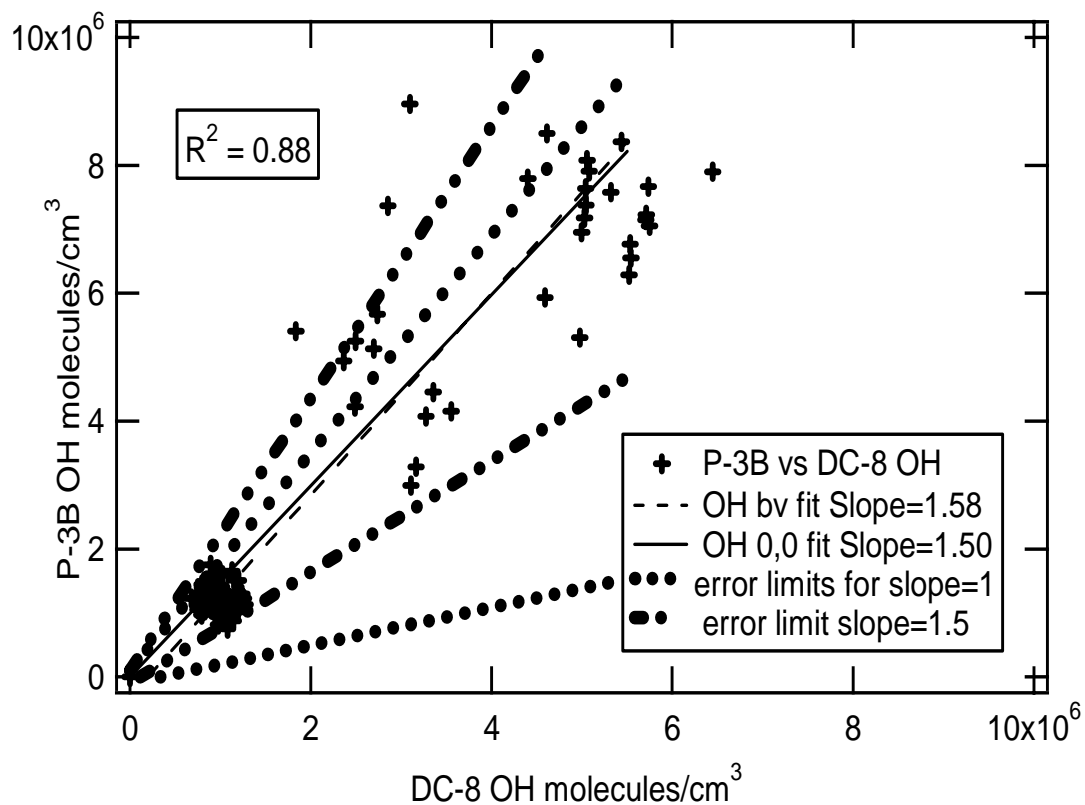


Fig 18

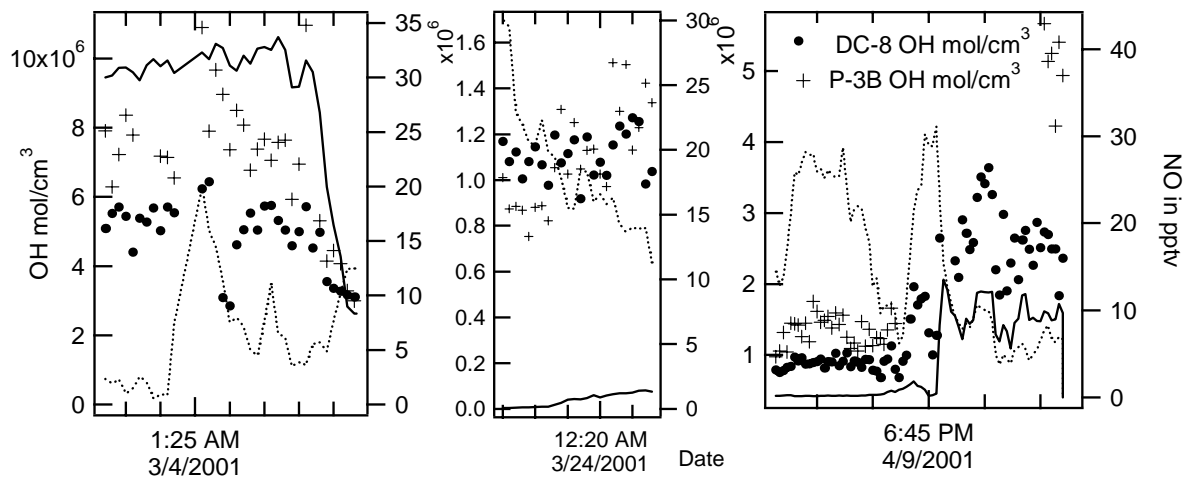


Fig 19

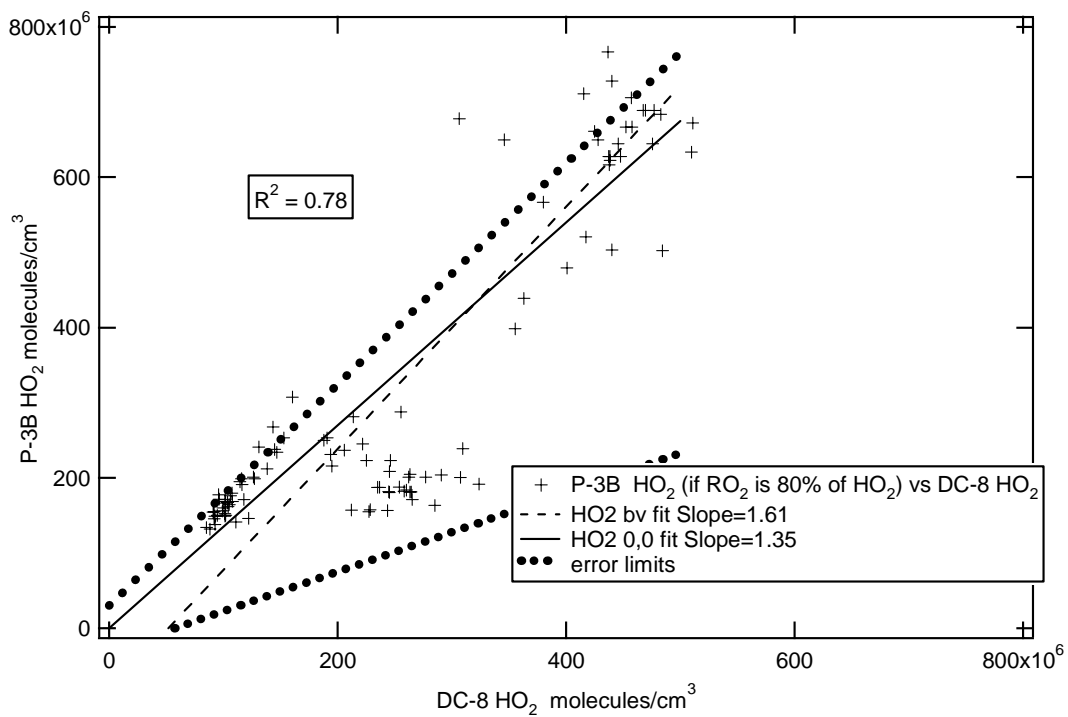


Fig 20

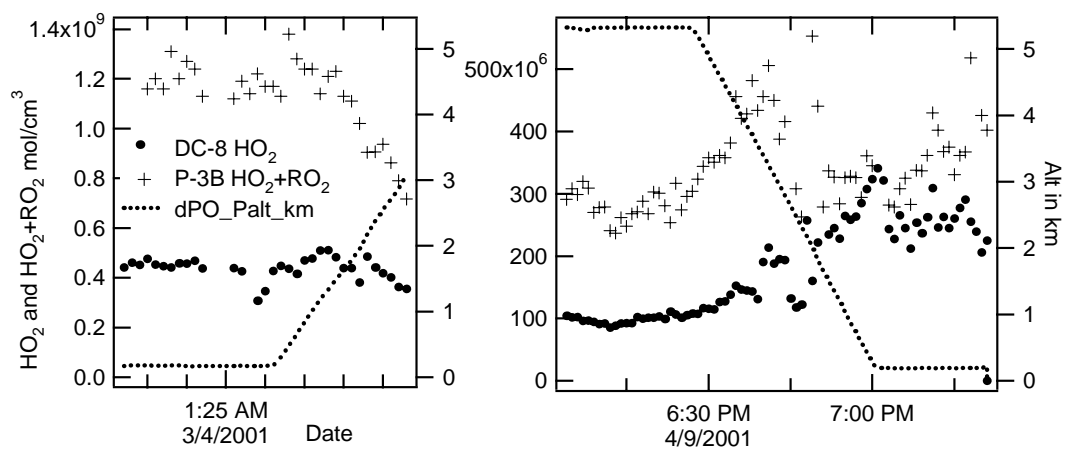


Fig 21

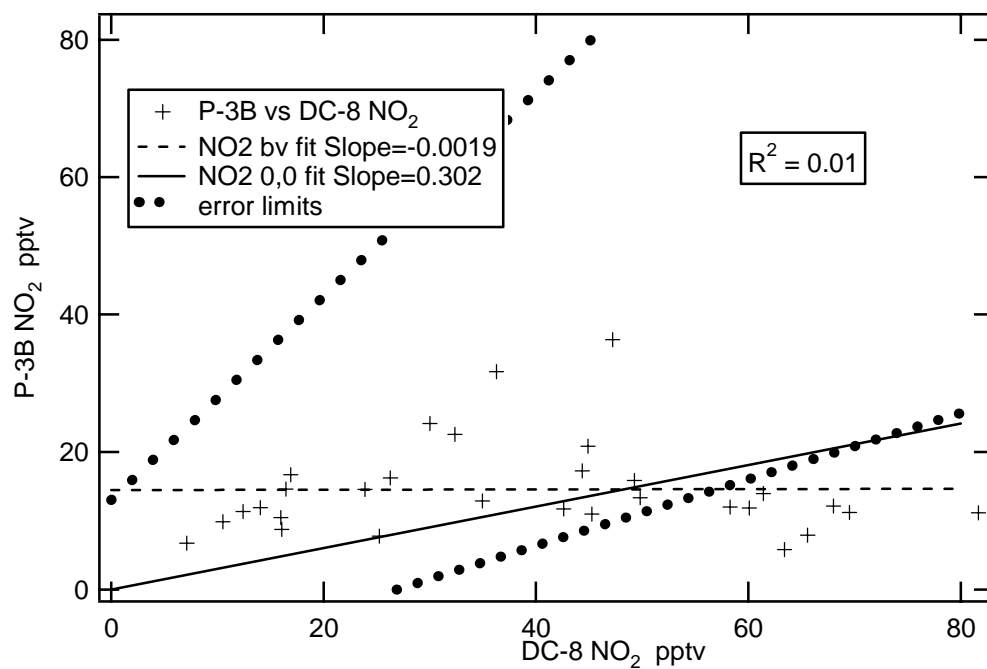


Fig 22

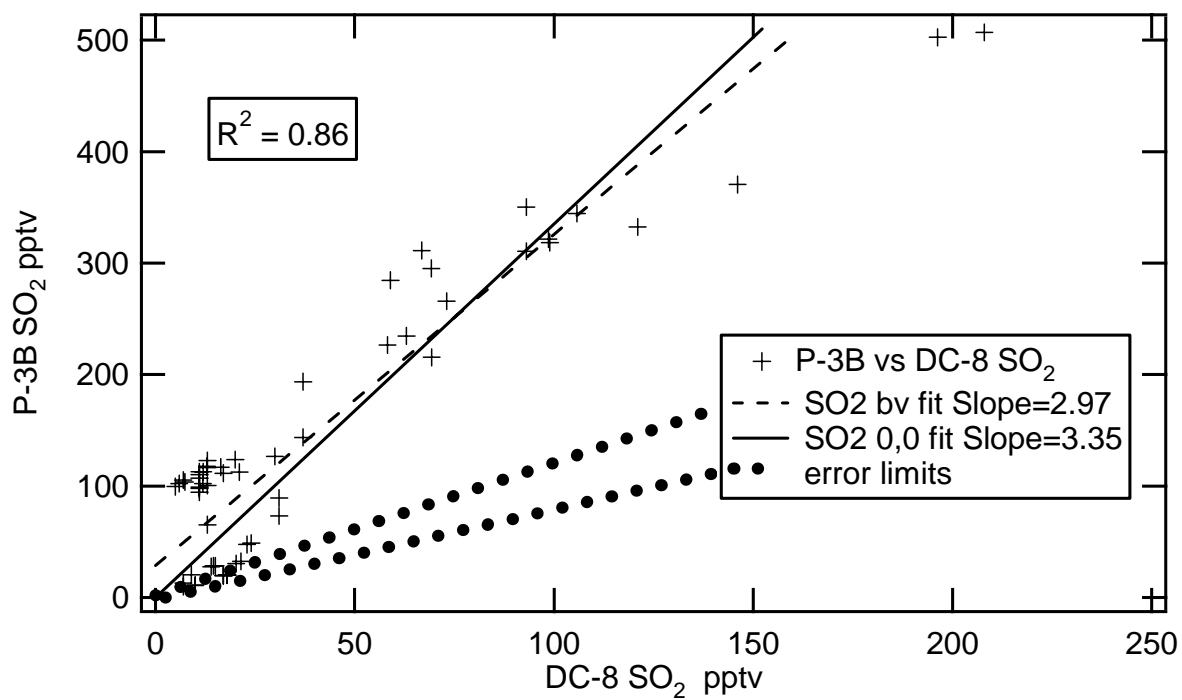


Fig 23

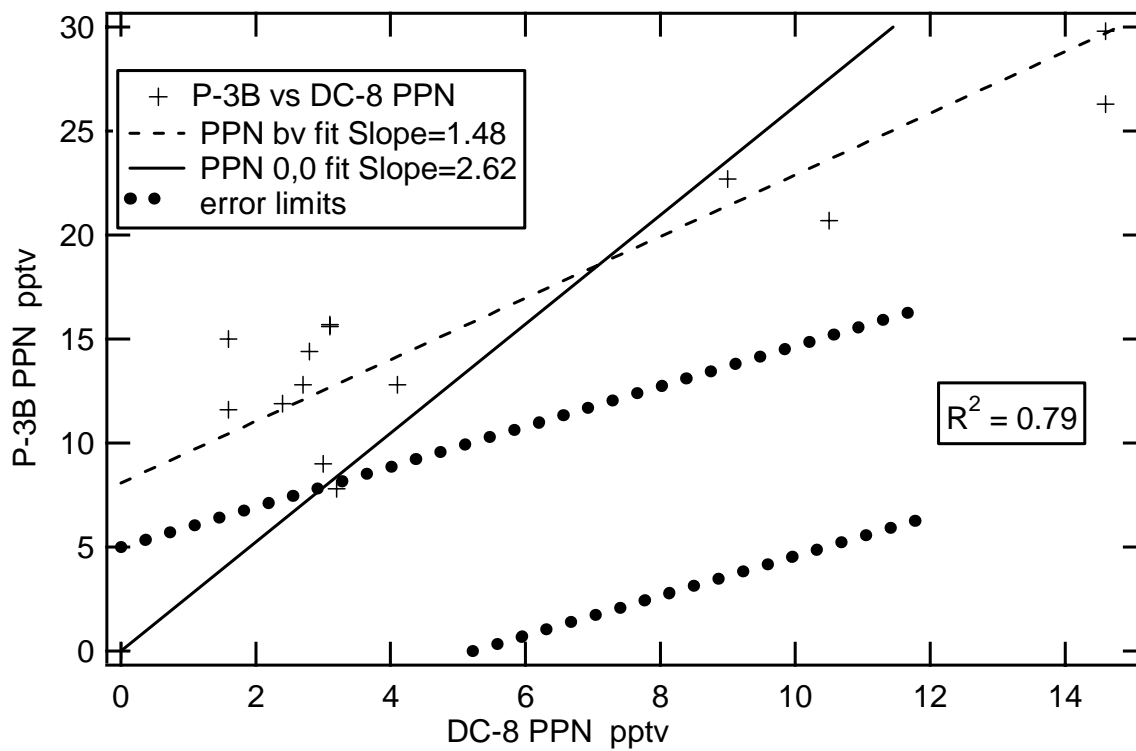


Fig 24

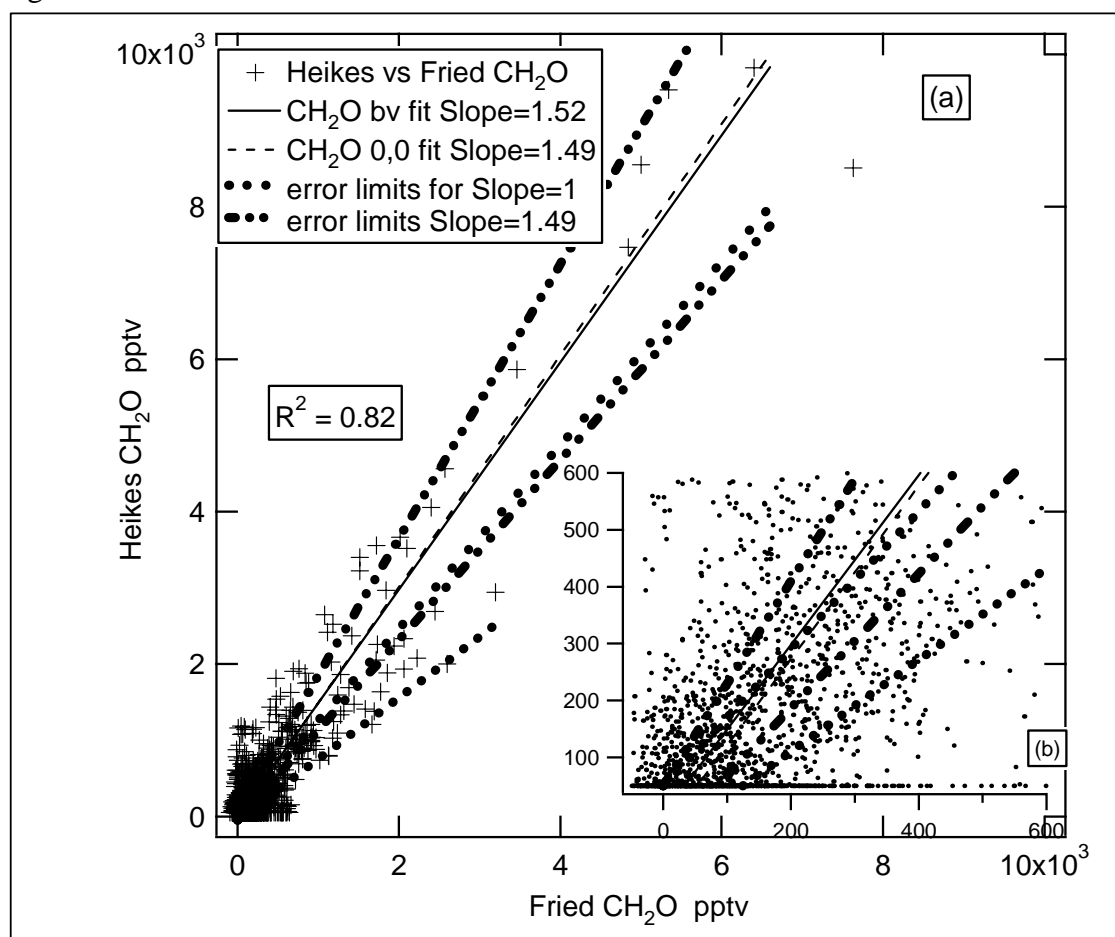


Fig 25

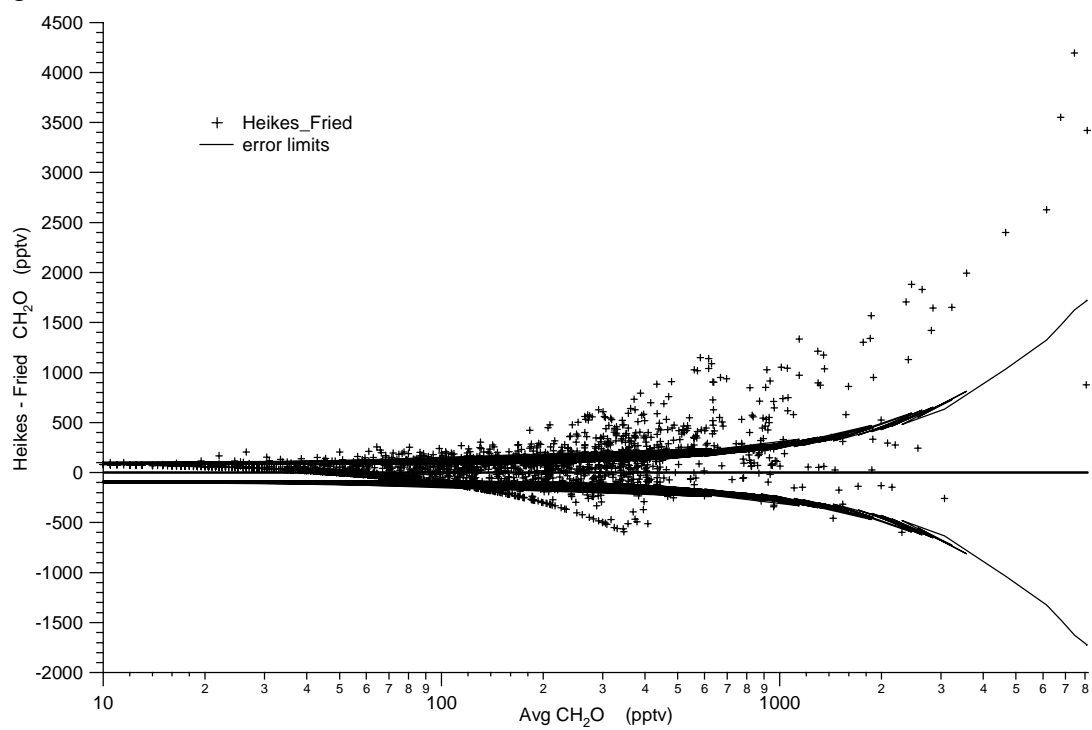


Fig 26

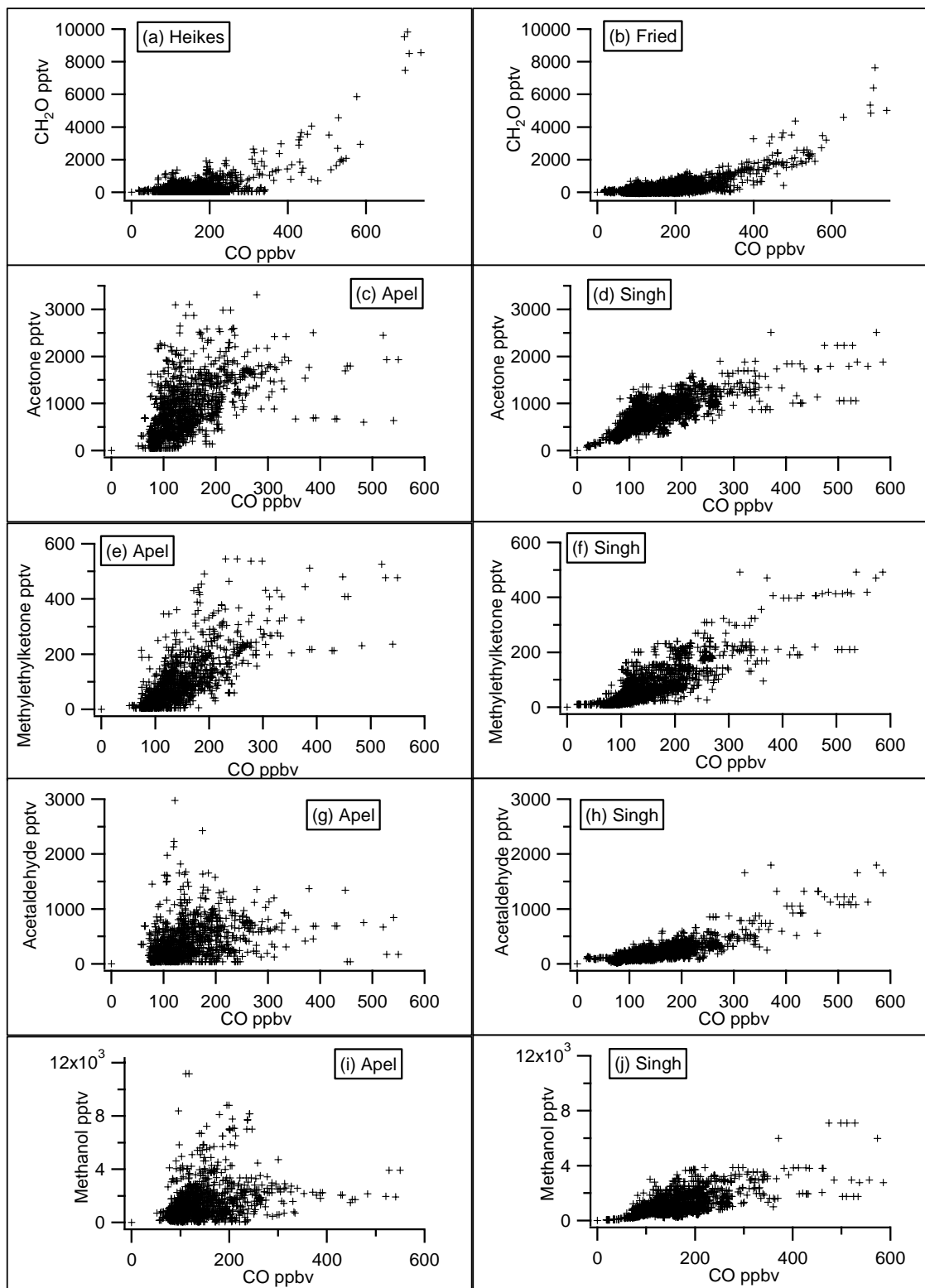


Fig 27

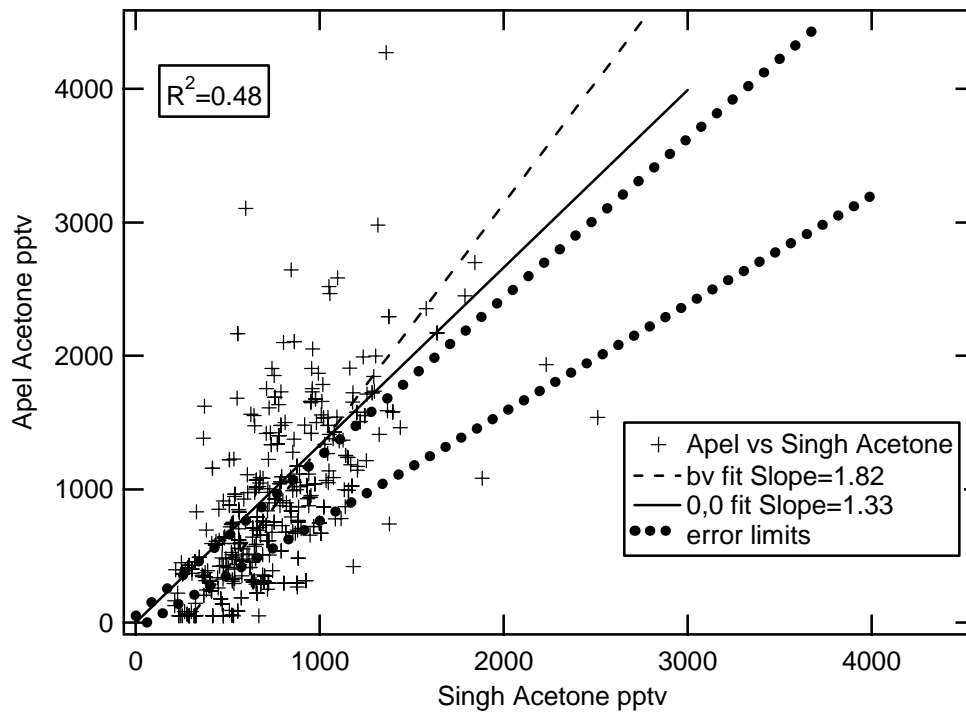


Fig 28

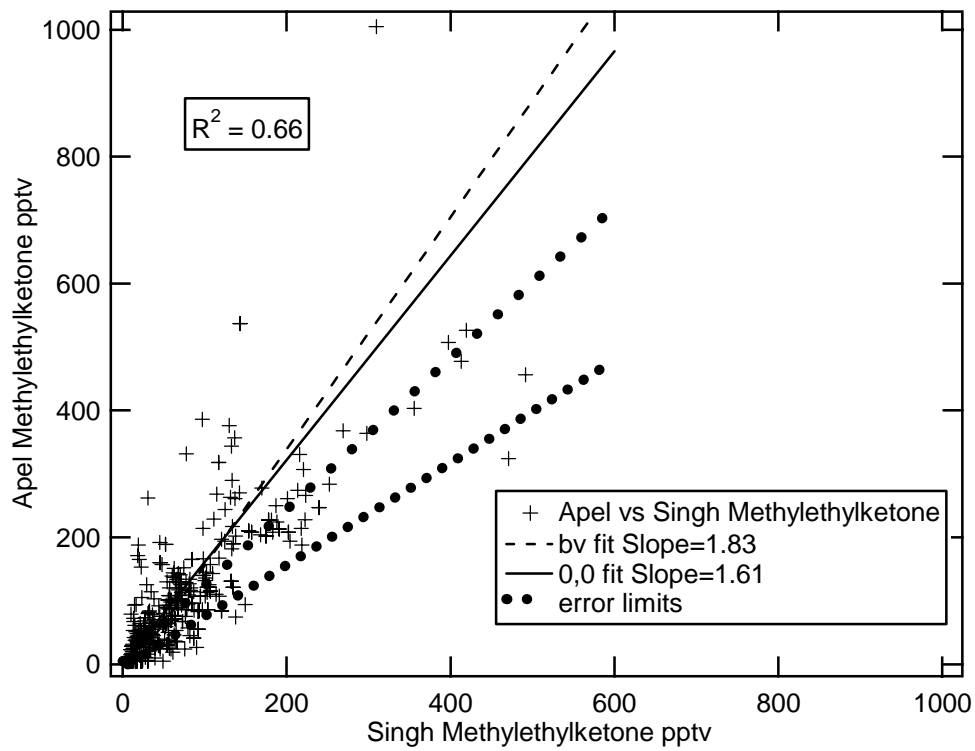


Fig 29

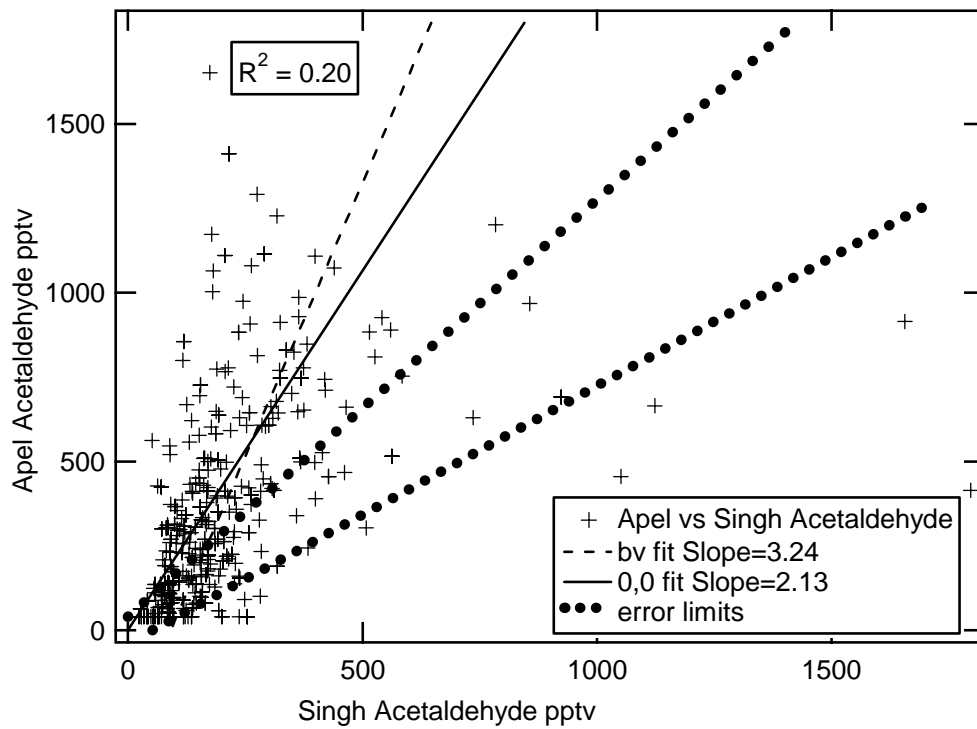


Fig 30

